

Tilburg University

## Asymptotic normality of maximum likelihood estimators obtained from normally distributed but dependent observations

Heijmans, R.D.H.; Magnus, J.R.

*Published in:*  
Econometric Theory

*Publication date:*  
1986

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Heijmans, R. D. H., & Magnus, J. R. (1986). Asymptotic normality of maximum likelihood estimators obtained from normally distributed but dependent observations. *Econometric Theory*, 2(3), 374-412.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# ASYMPTOTIC NORMALITY OF MAXIMUM LIKELIHOOD ESTIMATORS OBTAINED FROM NORMALLY DISTRIBUTED BUT DEPENDENT OBSERVATIONS

RISTO D. H. HEIJMANS

*University of Amsterdam*

JAN R. MAGNUS

*London School of Economics*

In this article we aim to establish intuitively appealing and verifiable conditions for the first-order efficiency and asymptotic normality of ML estimators in a multi-parameter framework, assuming joint normality but neither the independence nor the identical distribution of the observations. We present five theorems (and a large number of lemmas and propositions), each being a special case of its predecessor.

Tout le monde y croit cependant, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental.<sup>1</sup> (Poincaré [15])

## 1. INTRODUCTION

In econometric models the observations are, as a rule, not independent and identically distributed (i.i.d.). For example, the observations  $\{y_t\}$  in the linear regression model  $y_t = x_t'\beta + \varepsilon_t$  ( $t = 1, \dots, n$ ) are not i.i.d. even if the errors  $\{\varepsilon_t\}$  are, unless the regressors  $\{x_t\}$  are random and i.i.d., or  $x_t = c$  (a vector of constants). Most central limit theorems are, however, for the case where the observations are i.i.d., and hence not applicable to regression models.

A (very) preliminary version of this paper was presented at the 1979 European meeting of the Econometric Society in Athens. We thank the participants of workshops at LSE, Essex, Manchester, Nuffield College, ANU, Monash, and Sydney for their helpful comments. We are particularly grateful to Peter Robinson, B. B. van der Genugten, and the referees for their constructive remarks. Magnus' research was supported, in part, by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).



Among the useful exceptions we mention central limit theorems by Malinvaud [12, p. 250–253] for the case where the observations are independent but not identically distributed, Rozanov [17, p. 190–198] and Hannan [4, p. 220–229] for stationary stochastic processes, and Schöfeld [19] for  $m$ -dependent observations.

In searching for conditions which imply asymptotic normality of the maximum likelihood (ML) estimator of the parameters in nonlinear models, we again face the problem that the observations from which the ML estimator is obtained are not i.i.d. In Heijmans and Magnus [6] we studied this problem in a more general framework and established conditions which appear to be weaker and more readily applicable than usual. In particular, the regularity conditions never require convergences in probability to be uniform.

The present article is based on and extends from [6]. We consider a set  $y = (y_1, y_2, \dots, y_n)$  of observations, not necessarily independent or identically distributed, whose joint distribution is known to be normal,

$$y \simeq N(\mu(\gamma_o), \Omega(\gamma_o)), \quad (1)$$

where both the mean vector  $\mu$  and the covariance matrix  $\Omega$  may depend on an unknown parameter vector  $\gamma_o$  to be estimated. This set-up, apart from the assumed unconditional normality, is rather wide since it contains not only the nonlinear regression model with “fixed” regressors, but also linear models with lagged dependent observations, random regressors or random parameters. Notice that the covariance matrix of the observations may depend on parameters in the mean. We shall discuss the generality and limitations of our set-up more fully in Section 3.

The existence and consistency of ML estimators are assumed both in [6] and in the present article. We discuss both issues in detail in [7].

A brief review of the literature on ML estimation with generally dependent observations can be found in [6]. Here we only mention two predecessors of the present study. First, Anderson [1, p. 183–211], generalizing the important paper by Mann and Wald [13], considered the linear regression model with exogenous and lagged endogenous variables

$$y_t = \sum_{i=1}^k \beta_i x_{ti} + \sum_{j=1}^p \delta_j y_{t-j} + \varepsilon_t, \quad (t = 1, \dots, n), \quad (2)$$

where the  $\varepsilon_t$ 's are i.i.d.  $N(0, \sigma^2)$ . We note that Rubin [18] extended [13] to allow for exogenous variables much before Anderson, and that Schöfeld [19] gave the same extension as Anderson did. Many authors have improved upon Anderson's work in various ways. See for example, Hannan [5]. Secondly, Magnus [10] studied the linear regression model

$$y = X\beta_o + \varepsilon, \quad (3)$$



where  $X$  is nonrandom and of full rank,  $\varepsilon$  is distributed as  $N_n(0, \Omega(\theta_o))$ , and the parameters in  $\beta_o$  are functionally independent from those in  $\theta_o$ . Magnus' [10, Theorem 5] conditions go back to Weiss [20] and are very stringent. Both equations (2) and (3) are, of course, special cases of the model of equation (1).

This article contains five theorems, four propositions, and sixteen lemmas (of which seven are in Appendix A), and is organized as follows. In Section 2 we explain our notation, and state Theorem 1, which contains the extension of the central limit theorem to dependent observations previously obtained by us in [6]. In Section 3 we list the three basic assumptions; these are discussed briefly. The score vector, Hessian matrix, and Information matrix are derived in Section 4. In Section 5 we prove first- and second-order regularity of the loglikelihood function. Denoting by  $l_n(\gamma_o)$  the score vector (i.e., the derivative of the loglikelihood function) evaluated at the true value  $\gamma_o$ , we prove in Sections 6 and 7 some auxiliary results regarding the components of the random "difference" vector  $l_n(\gamma_o) - l_{n-1}(\gamma_o)$ , and show that the sequence  $\{l_n(\gamma_o)\}$  is a vector martingale.

All these results are finite results; the remainder of the article concerns asymptotic results. The first of these is the asymptotic normality of  $\{n^{-1/2}l_n(\gamma_o)\}$ , established in Section 8. Denoting the matrix of second-order partial derivatives of the loglikelihood function (the Hessian matrix) by  $R_n(\gamma)$ , we show in Section 9 that  $(1/n)R_n(\gamma_o)$  converges in probability to minus the Information matrix; in Section 10 conditions are given that guarantee sufficient continuity of  $(1/n)R_n(\gamma)$  that this function is in some sense close to  $(1/n)R_n(\gamma_o)$ , when  $\gamma$  is near  $\gamma_o$  and  $n$  is large. Our main result (Theorem 2) is stated in Section 11. A weaker version of Theorem 2 (Theorem 3), easier to apply in practice, is stated and discussed in Section 12. Section 13 contains the special, but important case where the covariance matrix and its derivatives up to the second order have locally bounded eigenvalues. As an example we work out (Section 14) the conditions for asymptotic normality in the case of first-order autocorrelation (Theorem 5) and compare these conditions with the literature. Two appendices containing the proofs conclude the article.

## 2. NOTATION AND SET-UP

The following notation is used. The defining equality is denoted  $:=$ , so that  $x := y$  defines  $x$  in terms of  $y$ .  $\mathbb{N} = \{1, 2, \dots\}$ , and  $\mathbb{R}^n$  denotes the  $n$ -dimensional Euclidean space. The eigenvalues of an  $n \times n$  matrix  $A$  are denoted  $\lambda_t(A)$ ,  $t = 1, \dots, n$ . To indicate the dimension of a vector or matrix, we often write  $\mu_{(n)} := (\mu_1, \mu_2, \dots, \mu_n)'$ , and  $\Omega_n$  for the  $n \times n$  matrix  $\Omega$ . If  $f(x)$  is an  $m \times 1$  vector function of an  $n \times 1$  vector  $x$ , then  $\partial f(x)/\partial x'$  denotes the  $m \times n$  matrix of partial derivatives, and  $\partial f(x)/\partial x_i$  the  $m \times 1$  vector of partial derivatives with respect to  $x_i$ ; if  $F(x)$  is an  $m \times p$  matrix function of  $x$ , then



$\partial F(x)/\partial x_i$  denotes the  $m \times p$  matrix of partial derivatives of  $F$  with respect to  $x_i$ .

The general set-up is as follows. (See also Heijmans and Magnus [6, Section 1.2].) Let  $\{y_1, y_2, \dots\}$  be a sequence of random variables, not necessarily independent or identically distributed. The joint density function of  $y_{(n)} := (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$  is denoted  $h_n(\cdot; \gamma_o)$ , and is assumed to be known except, of course, for  $\gamma_o$ , the true value of the parameter vector to be estimated. We assume that  $\gamma_o \in \overset{o}{\Gamma}$ , the interior of the parameter space  $\Gamma \subset \mathbb{R}^p$ . For every (fixed)  $y \in \mathbb{R}^n$  the real-valued function  $L_n(\gamma; y) := h_n(y; \gamma)$ ,  $\gamma \in \Gamma$ , is called the *likelihood* (function), and every value  $\hat{\gamma}_n(y) \in \Gamma$  with

$$L_n(\hat{\gamma}_n(y); y) = \sup_{\gamma \in \Gamma} L_n(\gamma; y) \quad (4)$$

is called an ML *estimate* of  $\gamma_o$ . Let  $M_n$  denote the set of  $y \in \mathbb{R}^n$  for which an ML estimate exists, i.e.,

$$M_n := \bigcup_{\gamma \in \Gamma} \left\{ y: y \in \mathbb{R}^n, L_n(\gamma; y) = \sup_{\phi \in \Gamma} L_n(\phi; y) \right\}, \quad n \in \mathbb{N}.$$

If there exists, for every  $n \in \mathbb{N}$ , a measurable function  $\hat{\gamma}_n$  from  $\mathbb{R}^n$  into  $\Gamma$  such that equation (4) holds for every  $y \in M_n$ , and a measurable subset  $M'_n$  of  $M_n$  such that  $P(M'_n) \rightarrow 1$  as  $n \rightarrow \infty$ , then we say that an ML *estimator*  $\{\hat{\gamma}_n(y_{(n)})\}$  of  $\gamma_o \in \Gamma$  exists *asymptotically almost surely* (a.a.s.).

All probabilities and expectations are taken with respect to the true underlying distribution. Thus we write  $P$  instead of  $P_{\gamma_o}$ ,  $E$  instead of  $E_{\gamma_o}$ , etc.

For fixed  $y \in \mathbb{R}^n$ ,  $\Lambda_n(\gamma) := \log L_n(\gamma; y)$  is the loglikelihood function and, if  $\Lambda_n$  is twice differentiable,  $l_n(\gamma) := \partial \Lambda_n(\gamma)/\partial \gamma$  denotes the  $p \times 1$  score vector and  $R_n(\gamma) := \partial^2 \Lambda_n(\gamma)/\partial \gamma \partial \gamma'$  the  $p \times p$  Hessian matrix with elements  $R_{nij}$ . We define  $L_o(\gamma) := 1$  and let  $g_n(\gamma) := L_n(\gamma)/L_{n-1}(\gamma)$  be the "conditional likelihood". Finally, we define  $\xi_{nj} := \partial \log g_n(\gamma_o)/\partial \gamma_j$  ( $j = 1, \dots, p$ ).

In Heijmans and Magnus [6] we considered the first-order efficiency and asymptotic normality of ML estimators obtained from generally dependent observations, assuming that the joint density of  $(y_1, \dots, y_n)$  is known (except, of course, for  $\gamma_o$ ), but without specifying this function. The following result, which will serve as our starting point, was proved there.

**THEOREM 1.** *Assume that*

1. *for every (fixed)  $n \in \mathbb{N}$  and  $y_{(n)} \in \mathbb{R}^n$ , the loglikelihood  $\Lambda_n(\gamma; y_{(n)})$  is twice continuously differentiable on  $\overset{o}{\Gamma}$ ;*
2.  $E\xi_{nj}^4 < \infty$  ( $n \in \mathbb{N}, j = 1, \dots, p$ );
3.  $E(\xi_{nj} | y_1, \dots, y_{n-1}) = 0$  a.s. ( $n \geq 2, j = 1, \dots, p$ );
4.  $p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \xi_{tj}^2 = 0$  ( $j = 1, \dots, p$ );



5.  $\lim_{n \rightarrow \infty} (1/n^2) \text{var} \sum_{t=1}^n E(\xi_{ti} \xi_{tj} | y_1, \dots, y_{t-1}) = 0 \quad (i, j = 1, \dots, p);$
6.  $\lim_{n \rightarrow \infty} (1/n^2) \sum_{t=1}^n \text{var} (\xi_{ti} \xi_{tj} - E(\xi_{ti} \xi_{tj} | y_1, \dots, y_{t-1})) = 0 \quad (i, j = 1, \dots, p);$
7. *there exists a finite positive definite  $p \times p$  matrix  $G_o$  such that*  

$$\lim_{n \rightarrow \infty} (1/n) E l_n(\gamma_o) l_n'(\gamma_o) = G_o;$$
8.  $p \lim_{n \rightarrow \infty} (1/n) R_n(\gamma_o) = -G_o;$
9. *for every  $\alpha > 0$  there exists a neighborhood  $N(\gamma_o)$  of  $\gamma_o$  such that for  $i, j = 1, \dots, p$ ,*

$$\lim_{n \rightarrow \infty} P \left[ (1/n) \sup_{\gamma \in N(\gamma_o)} |R_{nij}(\gamma) - R_{nij}(\gamma_o)| > \alpha \right] = 0.$$

Assume further that an ML estimator  $\{\hat{\gamma}_n\}$  of  $\gamma_o \in \overset{o}{\Gamma}$  exists asymptotically almost surely, and is weakly consistent. Then the sequence  $\{\hat{\gamma}_n\}$  is first-order efficient and asymptotically normally distributed, i.e.,

$$p \lim_{n \rightarrow \infty} (\sqrt{n}(\hat{\gamma}_n - \gamma_o) - (1/\sqrt{n})G_o^{-1}l_n(\gamma_o)) = 0$$

and

$$\sqrt{n}(\hat{\gamma}_n - \gamma_o) \xrightarrow{L} N(0, G_o^{-1}).$$

**Proof.** This is the special case of Heijmans and Magnus [6, Theorem 2] obtained by choosing  $j_n = i_n := E l_n'(\gamma_o) l_n(\gamma_o)$  and  $G_o = K_o$ , and assuming that  $i_n/n$  tends to a positive constant as  $n \rightarrow \infty$ . ■

In the remainder of this article we shall assume that the joint density is normal. The precise framework is described in the next section.

### 3. BASIC ASSUMPTIONS

The first assumption defines the structure which generates the observations.

**Assumption 1.** For every (fixed)  $n \in \mathbb{N}$ ,  $y_{(n)} := (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$  follows an  $n$ -variate normal distribution

$$y_{(n)} \simeq N(\mu_{(n)}(\gamma_o), \Omega_n(\gamma_o)), \quad \gamma_o \in \overset{o}{\Gamma} \subset \mathbb{R}^p,$$

where  $\gamma_o$  is the true (but unknown) value of the parameter vector to be estimated, and  $p$ , the dimension of  $\overset{o}{\Gamma}$ , is independent of  $n$ .

We further specify the functions involved in the next two assumptions.



Assumption 2.  $\mu_{(n)}: \Gamma \rightarrow \mathbb{R}^n$  and  $\Omega_n: \Gamma \rightarrow \mathbb{R}^{n \times n}_o$  are known twice continuously differentiable vector (matrix) functions on  $\Gamma$ , the interior of  $\Gamma$ , for every (fixed)  $n \in \mathbb{N}$ .

Assumption 3. The matrix  $\Omega_n(\gamma)$  is positively definite (hence nonsingular) for every  $n \in \mathbb{N}$  and  $\gamma \in \Gamma_o$ .

The assumption of unconditional normality is, of course, rather strong, but not as strong as it may seem. Let us consider the class of models included in this assumption. First, the classical nonlinear regression model

$$y_t = \phi(x_t, \beta_o) + \varepsilon_t, \quad (t = 1, \dots, n),$$

where  $\phi(\cdot)$  is the known response function,  $x_t$  is a nonrandom vector containing the values observed for each of the explanatory variables at time  $t$ , the  $\varepsilon_t$  are unobservable errors whose joint distribution is known to be

$$\varepsilon_{(n)} \simeq N(0, \Omega_n(\theta_o)),$$

and  $\beta_o$  and  $\theta_o$  are the true values of the parameter vectors to be estimated. Note, however, that our set-up allows for the fact that the covariance matrix of the observations may depend on parameters in the mean, thus including cases such as the type of heteroskedasticity where the variance of  $y_t$  is proportional to the square of the mean. Secondly, linear models with lagged dependent observations. For example,

$$y_t = \alpha_o + \beta_o y_{t-1} + \gamma_o x_t + \varepsilon_t, \quad t = 1, 2, \dots,$$

where  $\{x_t\}$  is a sequence of observations on the nonstochastic regressor,  $y_o$  is fixed, and  $\{\varepsilon_t\}$  is i.i.d.  $N(0, \sigma_o^2)$ . Then  $(y_1, y_2, \dots, y_n)$  is  $n$ -variate normally distributed, with

$$\mu_t := Ey_t = y_o \beta_o^t + \alpha_o \sum_{j=0}^{t-1} \beta_o^j + \gamma_o \sum_{j=0}^{t-1} \beta_o^j x_{t-j}, \quad t = 1, \dots, n$$

and

$$\omega_{st} := \text{cov}(y_s, y_t) = \sigma_o^2 \beta_o^{|s-t|} \sum_{j=0}^{\delta(s,t)} \beta_o^{2j}, \quad s, t = 1, \dots, n,$$

where  $\delta(s, t) := \min(s-1, t-1)$ . The covariance matrix  $\Omega = (\omega_{st})$  depends of course on  $\beta_o$ . The situation where the errors  $\{\varepsilon_t\}$  are not i.i.d. also falls within our framework; only the expression for  $\omega_{st}$  becomes more complicated. Thirdly, linear models with stochastically varying coefficients or stochastic regressors. The linearity, except in certain pathological situations, is essential in order to maintain unconditional normality. However, models that are linear in the stochastic regressors or lagged endogenous variables are allowed to be nonlinear in the (nonrandom) parameters, and models that are linear



in the stochastic parameters may be nonlinear in the (strictly exogenous) regressors.

The fact that the observations  $\{y_t\}$  are scalars is not restrictive, because of the assumed general dependence. The observations may be finite dimensional vectors and even the orders of these vectors may vary with  $t$ . The only requirement is that their joint distribution is multivariate normal.

Three further points are worth noting. First, the number of parameters is assumed to be fixed and independent of the number of observations. Secondly, the parameter space  $\Gamma$  is a subset of  $\mathbb{R}^p$ , the Euclidean space of (fixed) dimension  $p \geq 1$ , but not necessarily a  $p$ -dimensional interval. Thirdly, for fixed  $n$ , the covariance matrix  $\Omega_n(\gamma)$  (and therefore also  $\Omega_n^{-1}(\gamma)$ ) is required to be nonsingular for every value of  $\gamma \in \overset{o}{\Gamma}$ . But, for  $n \rightarrow \infty$ , there will in general be values of  $\gamma$  such that

$$|\Omega_n(\gamma)| \rightarrow 0 \quad \text{or} \quad |\Omega_n(\gamma)| \rightarrow \infty.$$

#### 4. THE DERIVATIVES OF THE LIKELIHOOD FUNCTION

The loglikelihood function based on  $n$  observations  $y_1, \dots, y_n$  is

$$\begin{aligned} \Lambda_n(\gamma) = & -(n/2) \log 2\pi - \frac{1}{2} \log |\Omega_n(\gamma)| \\ & - \frac{1}{2} (y_{(n)} - \mu_{(n)}(\gamma))' \Omega_n^{-1}(\gamma) (y_{(n)} - \mu_{(n)}(\gamma)), \quad (\gamma \in \Gamma). \end{aligned} \quad (5)$$

**LEMMA 1.** *Given Assumptions 1–3, the loglikelihood  $\Lambda_n(\gamma)$  is, for every (fixed)  $n \in \mathbb{N}$  and  $y_{(n)} \in \mathbb{R}^n$ , twice continuously differentiable on  $\overset{o}{\Gamma}$ .*

*Proof.* Obvious. ■

Lemma 1 ensures that the (log) likelihood is twice continuously differentiable in a neighborhood of the true parameter value  $\gamma_o$ .

Let us define the  $p \times 1$  score vector

$$l_n(\gamma) := \frac{\partial \Lambda_n(\gamma)}{\partial \gamma}, \quad (\gamma \in \overset{o}{\Gamma}), \quad (6)$$

with components  $l_{nj}(\gamma)$ ,  $j = 1, \dots, p$ ; the symmetric  $p \times p$  Hessian matrix

$$R_n(\gamma) := \frac{\partial^2 \Lambda_n(\gamma)}{\partial \gamma \partial \gamma'}, \quad (\gamma \in \overset{o}{\Gamma}), \quad (7)$$

with elements  $R_{nij}(\gamma)$ ,  $i, j = 1, \dots, p$ ; and the symmetric  $p \times p$  Information matrix<sup>2</sup>

$$G_n(\gamma_o) := -ER_n(\gamma_o), \quad (8)$$

with elements  $G_{nij}(\gamma_o)$ ,  $i, j = 1, \dots, p$ . We can now establish Proposition 1.



Proposition 1.<sup>3</sup> Given Assumptions 1–3, the typical elements of the  $p \times 1$  score vector  $l_n(\gamma)$ , the  $p \times p$  Hessian matrix  $R_n(\gamma)$ , and the  $p \times p$  Information matrix  $G_n(\gamma_o)$ , are given by

$$\begin{aligned}
 l_{nj}(\gamma) &= \frac{1}{2} \operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) + (y_{(n)} - \mu_{(n)}(\gamma))' \Omega_n^{-1}(\gamma) \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \\
 &\quad - \frac{1}{2} (y_{(n)} - \mu_{(n)}(\gamma))' \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} (y_{(n)} - \mu_{(n)}(\gamma)); \\
 R_{nij}(\gamma) &= - \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma) \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \\
 &\quad - \frac{1}{2} \operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) \\
 &\quad + \frac{1}{2} \operatorname{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right) + (y_{(n)} - \mu_{(n)}(\gamma))' q_{nij}(\gamma) \\
 &\quad - \frac{1}{2} (y_{(n)} - \mu_{(n)}(\gamma))' \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} (y_{(n)} - \mu_{(n)}(\gamma)),
 \end{aligned}$$

where

$$q_{nij}(\gamma) := \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} + \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} + \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j};$$

and

$$\begin{aligned}
 G_{nij}(\gamma_o) &= \left( \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma_o) \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma_j} \\
 &\quad + \frac{1}{2} \operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma_o)}{\partial \gamma_i} \Omega_n(\gamma_o) \frac{\partial \Omega_n^{-1}(\gamma_o)}{\partial \gamma_j} \Omega_n(\gamma_o) \right).
 \end{aligned}$$

From Proposition 1 we easily obtain a compact expression for the Information matrix (rather than for a typical element), namely<sup>4</sup>

$$\begin{aligned}
 G_n(\gamma_o) &= \left( \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma'} \right)' \Omega_n^{-1}(\gamma_o) \left( \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma'} \right) \\
 &\quad + \frac{1}{2} \left( \frac{\partial \operatorname{vec} \Omega_n^{-1}(\gamma_o)}{\partial \gamma'} \right)' (\Omega_n(\gamma_o) \otimes \Omega_n(\gamma_o)) \left( \frac{\partial \operatorname{vec} \Omega_n^{-1}(\gamma_o)}{\partial \gamma'} \right), \tag{10}
 \end{aligned}$$

which shows that the Information matrix is positive semidefinite for every  $n \in \mathbb{N}$ .

For the special case where  $\mu$  is a function of “structural” parameters  $\beta$ ,  $\Omega$  is a function of “covariance” parameters  $\theta$ , and  $\beta$  and  $\theta$  are functionally



unrelated, we note that the Information matrix is block-diagonal (see also Magnus [10, Theorem 3]):

$$G_n(\beta_o, \theta_o) = \begin{pmatrix} G_n^\beta(\beta_o, \theta_o) & 0 \\ 0 & \frac{1}{2}G_n^\theta(\theta_o) \end{pmatrix}, \quad (11)$$

where

$$G_n^\beta(\beta, \theta) = \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta'} \right)' \Omega_n^{-1}(\theta) \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta'} \right) \quad (12)$$

and  $G_n^\theta(\theta)$  is a symmetric matrix with typical element

$$(G_n^\theta(\theta))_{ij} = \text{tr} \left( \frac{\partial \Omega_n^{-1}(\theta)}{\partial \theta_i} \Omega_n(\theta) \frac{\partial \Omega_n^{-1}(\theta)}{\partial \theta_j} \Omega_n(\theta) \right). \quad (13)$$

## 5. REGULARITY

Our basic assumptions also suffice to prove first- and second- order regularity.

Proposition 2.<sup>5</sup> Given Assumptions 1–3, the loglikelihood function  $\Lambda_n(\gamma)$  is regular with respect to its first and second derivatives, i.e.,

$$El_n(\gamma_o) = 0, \quad -ER_n(\gamma_o) = El_n(\gamma_o)l'_n(\gamma_o).$$

## 6. FINITE PROPERTIES OF $\xi_{tj}$

Let  $g_t(\gamma) := L_t(\gamma)/L_{t-1}(\gamma)$  denote the “conditional likelihood” function, and let  $L_o(\gamma) := 1$ . In this section we are interested in the partial derivatives of  $\log g_t(\gamma)$  evaluated at the true value  $\gamma_o$ , i.e.,  $\xi_{tj} := \frac{\partial}{\partial \gamma_j} \log g_t(\gamma_o)$ ,  $j = 1, \dots, p$ .

First, however, we recall some properties of the Cholesky decomposition.

Given any positive definite  $n \times n$  matrix  $\Omega_n$ , there exists a unique diagonal  $n \times n$  matrix  $A_n$  with positive diagonal elements, and a unique unit upper triangular<sup>6</sup>  $n \times n$  matrix  $Z_n$ , such that

$$\Omega_n^{-1} = Z_n A_n^{-1} Z_n'. \quad (14)$$

This is the *Cholesky decomposition* of  $\Omega_n^{-1}$ . The diagonal elements  $\alpha_1, \dots, \alpha_n$  of  $A_n$  are called the *Cholesky numbers*. The first  $t$  components of the  $t$ -th column of  $Z_n$  form a  $t \times 1$  vector, which we denote  $z_t$  ( $t = 1, \dots, n$ ). Notice that the last ( $t$ -th) component of  $z_t$  is 1. The vectors  $z_1, \dots, z_n$  are called the *Cholesky vectors*.



Now let  $\Omega_n$  be the  $n \times n$  upper-left submatrix of a positive definite  $(n+1) \times (n+1)$  matrix  $\Omega_{n+1}$ . The matrix  $\Omega_{n+1}$  possesses  $n+1$  Cholesky numbers (vectors). A crucial property of the Cholesky decomposition is that the first  $n$  Cholesky numbers (vectors) of  $\Omega_{n+1}$  are precisely the Cholesky numbers (vectors) of  $\Omega_n$ . Hence,  $\alpha_t$  (and  $z_t$ ) unambiguously denotes the  $t$ -th Cholesky number (vector), *irrespective of the value of  $n$* .

The following two properties are easily established:

$$|\Omega_n| = \prod_{t=1}^n \alpha_t, \quad (15)$$

and

$$c'_{(n)} \Omega_n^{-1} c_{(n)} = \sum_{t=1}^n (\alpha_t^{-1/2} z'_t c_{(t)})^2, \quad (16)$$

given any set of real numbers  $c_1, \dots, c_n$ , where  $c_{(t)} := (c_1, \dots, c_t)'$ ,  $t = 1, \dots, n$ .

Using equations (15) and (16), we rewrite the loglikelihood function  $\Lambda_n$ , given in equation (5), as

$$\begin{aligned} \Lambda_n(\gamma) &= -(n/2) \log 2\pi - \frac{1}{2} \log |\Omega_n(\gamma)| \\ &\quad - \frac{1}{2} (y_{(n)} - \mu_{(n)}(\gamma))' \Omega_n^{-1}(\gamma) (y_{(n)} - \mu_{(n)}(\gamma)) \\ &= -(n/2) \log 2\pi - \frac{1}{2} \log \prod_{t=1}^n \alpha_t(\gamma) \\ &\quad - \frac{1}{2} \sum_{t=1}^n (\alpha_t^{-1/2}(\gamma) z'_t(\gamma) (y_{(t)} - \mu_{(t)}(\gamma)))^2 \\ &= -\frac{1}{2} \sum_{t=1}^n (\log 2\pi \alpha_t(\gamma) + (\alpha_t^{-1/2}(\gamma) z'_t(\gamma) (y_{(t)} - \mu_{(t)}(\gamma))))^2. \end{aligned}$$

Hence,

$$\log g_t(\gamma) = -\frac{1}{2} \log 2\pi \alpha_t(\gamma) - \frac{1}{2} (\alpha_t^{-1/2}(\gamma) z'_t(\gamma) (y_{(t)} - \mu_{(t)}(\gamma)))^2, \quad (17)$$

and its first-order partial derivatives  $\xi_{tj} := \partial \log g_t(\gamma_o) / \partial \gamma_j$  evaluated at  $\gamma_o$  can be conveniently expressed as

$$\xi_{tj} = -\frac{1}{2} \alpha_{tj} (1 - u_t^2) + (\delta_{tj} - w_{tj}) u_t, \quad (18)$$

where

$$\alpha_{tj} := \partial \log \alpha_t(\gamma_o) / \partial \gamma_j, \quad (19)$$

$$\delta_{tj} := (1/\sqrt{\alpha_t(\gamma_o)}) (\partial \mu_{(t)}(\gamma_o) / \partial \gamma_j)' z_t(\gamma_o), \quad (20)$$

$$u_t := (1/\sqrt{\alpha_t(\gamma_o)}) z'_t(\gamma_o) \varepsilon_{(t)}, \quad (21)$$



$$w_{tj} := (1/\sqrt{\alpha_t(\gamma_o)}) (\partial z_t(\gamma_o)/\partial \gamma_j)' \varepsilon_{(t)}, \quad (22)$$

$$\varepsilon_{(t)} := y_{(t)} - \mu_{(t)}(\gamma_o). \quad (23)$$

We note three important features of  $u_t$  and  $w_{tj}$  which render equation (18) for  $\xi_{tj}$  particularly useful. First, since  $\varepsilon_{(n)} \simeq N(0, \Omega_n(\gamma_o))$  and  $\Omega^{-1} = ZA^{-1}Z'$ , we have

$$A_n^{-1/2}(\gamma_o)Z'_n(\gamma_o)\varepsilon_{(n)} \simeq N(0, I_n).$$

Hence,  $u_1, u_2, \dots$  are independent and identically distributed as  $N(0, 1)$ .

Secondly, if  $e_k$  ( $k = 1, \dots, t$ ) denotes the  $k$ -th column of  $I_t$ , and  $\varepsilon_k$  the  $k$ -th component of  $\varepsilon_{(t)}$ , then, for  $k = 1, \dots, t-1$ ,

$$\begin{aligned} E\varepsilon_k u_t &= E(e'_k \varepsilon_{(t)}) (\alpha_t^{-1/2}(\gamma_o) \varepsilon'_{(t)} z_t(\gamma_o)) = \alpha_t^{-1/2}(\gamma_o) e'_k \Omega_t(\gamma_o) z_t(\gamma_o) \\ &= \alpha_t^{-1/2}(\gamma_o) e'_k (\alpha_t(\gamma_o) e_t) = \alpha_t^{1/2}(\gamma_o) e'_k e_t = 0, \end{aligned}$$

because

$$\begin{aligned} \Omega_t z_t &= (Z'_t)^{-1} A_t Z_t^{-1} Z_t e_t = (Z'_t)^{-1} A_t e_t \\ &= \alpha_t (Z'_t)^{-1} e_t = \alpha_t e_t. \end{aligned}$$

Hence,  $u_t$  is independent from  $\varepsilon_1, \dots, \varepsilon_{t-1}$ .

Thirdly, since  $w_{tj}$  depends only on  $\varepsilon_{(t-1)}$  and not on  $\varepsilon_t$  (because the last component of  $\partial z_t(\gamma_o)/\partial \gamma_j$  is zero), we also have that  $u_t$  and  $w_{tj}$  are independent.

Let us now establish some properties of  $\xi_{tj}$ .

**LEMMA 2.** *Given Assumptions 1–3,  $E\xi_{nj}^4 < \infty$  for every  $n \in \mathbb{N}$  and  $j = 1, \dots, p$ .*

*Proof.* Obvious. ■

It is also easy to prove that

$$E(\xi_{tj} | \varepsilon_{(t-1)}) = 0 \quad \text{a.s.} \quad (24)$$

Further, we have

$$\begin{aligned} \xi_{ti} \xi_{tj} &= \frac{1}{4} \alpha_{ti} \alpha_{tj} (1 - u_t^2)^2 + (\delta_{ti} - w_{ti})(\delta_{tj} - w_{tj}) u_t^2 \\ &= -\frac{1}{2} ((\delta_{ti} - w_{ti}) \alpha_{tj} + (\delta_{tj} - w_{tj}) \alpha_{ti}) (1 - u_t^2) u_t, \end{aligned}$$

and hence

$$E(\xi_{ti} \xi_{tj} | \varepsilon_{(t-1)}) = \frac{1}{2} \alpha_{ti} \alpha_{tj} + (\delta_{ti} - w_{ti})(\delta_{tj} - w_{tj}). \quad (25)$$



From equations (24) and (25) we easily obtain

$$\begin{aligned}\sigma_{ij}^2 &:= \text{var } \xi_{ij} = E\xi_{ij}^2 = EE(\xi_{ij}^2 | \varepsilon_{(t-1)}) \\ &= \frac{1}{2}\alpha_{ij}^2 + \delta_{ij}^2 + (1/\alpha_t(\gamma_o))(\partial z_t(\gamma_o)/\partial \gamma_j)' \Omega_t(\gamma_o)(\partial z_t(\gamma_o)/\partial \gamma_j).\end{aligned}\quad (26)$$

Finally, let us consider

$$\begin{aligned}v_{tij} &:= \xi_{ti}\xi_{tj} - E(\xi_{ti}\xi_{tj} | \varepsilon_{(t-1)}) \\ &= \frac{1}{4}\alpha_{ti}\alpha_{tj}(u_t^4 - 2u_t^2 - 1) - (\delta_{ti} - w_{ti})(\delta_{tj} - w_{tj})(1 - u_t^2) \\ &\quad - \frac{1}{2}((\delta_{ti} - w_{ti})\alpha_{tj} + (\delta_{tj} - w_{tj})\alpha_{ti})(1 - u_t^2)u_t.\end{aligned}$$

Using  $Eu_t^6 = 15$  and  $Eu_t^8 = 105$ , we have

$$\begin{aligned}2E(v_{tij}^2 | \varepsilon_{(t-1)}) &= 7\alpha_{ti}^2\alpha_{tj}^2 + 4(\delta_{ti} - w_{ti})^2(\delta_{tj} - w_{tj})^2 \\ &\quad + 5((\delta_{ti} - w_{ti})\alpha_{tj} + (\delta_{tj} - w_{tj})\alpha_{ti})^2 \\ &\quad + 4\alpha_{ti}\alpha_{tj}(\delta_{ti} - w_{ti})(\delta_{tj} - w_{tj}),\end{aligned}$$

and hence, after some elementary calculations,

$$\text{var } v_{tij} = EE(v_{tij}^2 | \varepsilon_{(t-1)}) \leq 14\sigma_{ti}^2\sigma_{tj}^2. \quad (27)$$

## 7. $\{l_n(\gamma_o)\}$ IS A VECTOR MARTINGALE

We shall now show that  $l_n(\gamma_o)$ , i.e., the derivative of the loglikelihood function evaluated at  $\gamma_o$ , is a vector martingale. (See Heijmans and Magnus [6, Section 3] for a definition and some properties of vector martingales.) The relevance of this fact lies in the possibility to use (vector) martingale limit theory to prove the asymptotic normality of  $\{n^{-1/2}l_n(\gamma_o)\}$ . We first establish Lemma 3.

**LEMMA 3.** *Given Assumptions 1–3,*

$$E(\xi_{nj} | y_1, \dots, y_{n-1}) = 0 \quad \text{a.s.} \quad (n \geq 2, j = 1, \dots, p).$$

*Proof.* This follows from equation (24). ■

An immediate consequence of Lemma 3 is Proposition 3.

**Proposition 3.** *Given Assumptions 1–3, the sequence  $\{l_n(\gamma_o), n \in \mathbb{N}\}$  is a vector martingale.*

*Proof.* Since  $l_{nj}(\gamma_o) = \sum_{t=1}^n \xi_{tj}$ , Lemma 3 implies that  $E(l_{nj}(\gamma_o) | y_1, \dots, y_{n-1}) = l_{n-1,j}(\gamma_o)$ . Further  $l_n(\gamma_o)$  is a continuous (hence measurable) function of  $y_1, \dots, y_n$  with finite (in fact, zero) expectation. Hence, the result follows from Lemma 1 of Heijmans and Magnus [6]. ■



## 8. ASYMPTOTIC NORMALITY OF $\{n^{-1/2}l_n(\gamma_o)\}$

So far, all our results were finite results. From this section onwards, all results are asymptotic results. The first asymptotic result is the asymptotic normality of  $\{n^{-1/2}l_n(\gamma_o)\}$ , for which we make two further assumptions.

Assumption 4. If  $G_n(\gamma)$ ,  $\gamma \in \overset{o}{\Gamma}$ , denotes the positive semidefinite  $p \times p$  matrix whose  $ij$ -th element is given by

$$G_{nij}(\gamma) := \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) + \frac{1}{2} \operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right),$$

then, for all  $\gamma \in \overset{o}{\Gamma}$ , the matrix  $(1/n)G_n(\gamma)$  converges, as  $n \rightarrow \infty$ , to a positive definite  $p \times p$  matrix  $G_o(\gamma)$ .

Assumption 5. In terms of the (unique) Cholesky decomposition of  $\Omega_n^{-1}$ :  $\Omega_n^{-1} = Z_n A_n^{-1} Z_n'$ , where  $A_n$  is a diagonal  $n \times n$  matrix with positive diagonal elements and  $Z_n$  is a unit upper triangular  $n \times n$  matrix, we have, for every  $\gamma \in \overset{o}{\Gamma}$ ,

$$\lim_{n \rightarrow \infty} (1/n^2) \operatorname{tr} \left( \frac{\partial Z_n(\gamma)}{\partial \gamma_i} A_n^{-1}(\gamma) \left( \frac{\partial Z_n(\gamma)}{\partial \gamma_i} \right)' \Omega_n(\gamma) \right)^2 = 0, \quad (i = 1, \dots, p).$$

Assumption 4 ensures that the Information matrix  $G_n(\gamma_o)$ , which is positive semidefinite for every  $n \in \mathbb{N}$  (see section 4), is in fact positive definite for  $n$  sufficiently large. In the linear regression model  $y = X\beta_o + \varepsilon$ , with  $\varepsilon \simeq N(0, \sigma_o^2 V)$ ,  $V$  known, we commonly require that  $(1/n)X'V^{-1}X$  tends to a positive definite matrix as  $n \rightarrow \infty$ . Assumption 4 is the equivalent condition for our case.

Assumption 5 requires that  $(1/n^2) \operatorname{tr}(C_{ni}^2(\gamma)) \rightarrow 0$  as  $n \rightarrow \infty$ , where

$$C_{ni}(\gamma) := \Omega_n^{1/2}(\gamma) \left( \frac{\partial Z_n(\gamma)}{\partial \gamma_i} \right) A_n^{-1}(\gamma) \left( \frac{\partial Z_n(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{1/2}(\gamma).$$

But since

$$\operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \right)^2 = 2 \operatorname{tr}(C_{ni}(\gamma)) + \operatorname{tr} \left( A_n^{-1}(\gamma) \frac{\partial A_n(\gamma)}{\partial \gamma_i} \right)^2,$$

we obtain

$$\operatorname{tr}(C_{ni}(\gamma)) \leq G_{nii}(\gamma),$$



and so

$$\begin{aligned} (1/n^2)\mu_{ni}^2(\gamma) &\leq (1/n^2)\text{tr}(C_{ni}^2(\gamma)) \leq ((1/n)\mu_{ni}(\gamma))((1/n)\text{tr}(C_{ni}(\gamma))) \\ &\leq ((1/n)\mu_{ni}(\gamma))((1/n)G_{nii}(\gamma)), \end{aligned}$$

where

$$\mu_{ni}(\gamma) := \max_{1 \leq t \leq n} \lambda_t(C_{ni}(\gamma)).$$

Hence, given Assumption 4, Assumption 5 is equivalent to

$$\lim_{n \rightarrow \infty} (1/n)\mu_{ni}(\gamma) = 0, \quad (i = 1, \dots, p).$$

This shows that Assumption 5 is a weak assumption. For, knowing that the average of the  $n$  eigenvalues of  $C_{ni}(\gamma)$  is bounded (for every fixed  $\gamma \in \Gamma$ ), we do not require that each eigenvalue of  $C_{ni}(\gamma)$  is bounded, but only that each eigenvalue is  $o(n)$ .

The following three lemmas can now be established.

LEMMA 4. *Given Assumptions 1–4,*

$$p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \xi_{tj}^2 = 0, \quad (j = 1, \dots, p).$$

LEMMA 5. *Given Assumptions 1–5,*

$$\lim_{n \rightarrow \infty} (1/n^2) \text{var} \sum_{t=1}^n E(\xi_{ti}\xi_{tj} | y_1, \dots, y_{t-1}) = 0, \quad (i, j = 1, \dots, p).$$

LEMMA 6. *Given Assumptions 1–4,*

$$\begin{aligned} \lim_{n \rightarrow \infty} (1/n^2) \sum_{t=1}^n \text{var}(\xi_{ti}\xi_{tj} - E(\xi_{ti}\xi_{tj} | y_1, \dots, y_{t-1})) &= 0, \\ (i, j = 1, \dots, p). \end{aligned}$$

We now have all the ingredients to prove the asymptotic normality of the sequence  $\{n^{-1/2}l_n(\gamma_o)\}$ .

Proposition 4. *Given Assumptions 1–5, the sequence  $\{n^{-1/2}l_n(\gamma_o), n \in \mathbb{N}\}$  is asymptotically normally distributed, i.e.,*

$$n^{-1/2}l_n(\gamma_o) \xrightarrow{L} N(0, G_o(\gamma_o)).$$



## 9. CONVERGENCE (IN PROBABILITY) OF THE HESSIAN MATRIX

In this section we shall prove that the Hessian matrix  $R_n(\gamma_o)$  divided by  $n$  converges in probability to minus the asymptotic Information matrix  $G_o(\gamma_o)$ . First, however, we establish Lemma 7, which we shall need later in the proof of Theorem 2.

LEMMA 7. *Given Assumptions 1–4,*

$$\lim_{n \rightarrow \infty} (1/n)El_n(\gamma_o)l'_n(\gamma_o) = G_o(\gamma_o).$$

Proof. We have  $(1/n)El_n(\gamma_o)l'_n(\gamma_o) = -(1/n)ER_n(\gamma_o) = (1/n)G_n(\gamma_o) \rightarrow G_o(\gamma_o)$ , as  $n \rightarrow \infty$ , using second-order regularity (Proposition 2), Definition (8) and Assumption 4. ■

Let us now consider the convergence of  $(1/n)R_n(\gamma_o)$  to  $-G_o(\gamma_o)$ . We shall make the following two additional assumptions.

Assumption 6. For every  $\gamma \in \overset{o}{\Gamma}$ ,

$$\lim_{n \rightarrow \infty} (1/n^2) \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right)^2 = 0, \quad (i, j = 1, \dots, p).$$

Assumption 7. For every  $\gamma \in \overset{o}{\Gamma}$ ,

$$(i) \quad \lim_{n \rightarrow \infty} (1/n^2) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) = 0, \\ (i, j = 1, \dots, p)$$

and

$$(ii) \quad \lim_{n \rightarrow \infty} (1/n^2) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) = 0, \quad (i, j = 1, \dots, p).$$

Assumption 6 looks similar to Assumption 5, but neither is implied by the other. A necessary condition for Assumption 6 to hold is that

$$(1/n) \max_{1 \leq t \leq n} \left| \lambda_t \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right) \right| \rightarrow 0$$



as  $n \rightarrow \infty$ ; sufficient is that

$$(1/\sqrt{n}) \max_{1 \leq i \leq n} \left| \lambda_i \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right) \right| \rightarrow 0$$

as  $n \rightarrow \infty$ .

Assumption 7 is implied by the following conditions:

$$(i) \quad (1/n) \max_{1 \leq i \leq n} \lambda_i(\Omega_n(\gamma)) \rightarrow 0,$$

as  $n \rightarrow \infty$ , for every  $\gamma \in \overset{o}{\Gamma}$ , and

$$(ii) \quad \text{there exists a continuous function } M: \overset{o}{\Gamma} \rightarrow \mathbb{R} \text{ such that for all } n \in \mathbb{N} \text{ and } \gamma \in \overset{o}{\Gamma},$$

$$(1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \right)^2 \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M(\gamma), \quad (i, j = 1, \dots, p),$$

and

$$(1/n) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)' \Omega_n^{-2}(\gamma) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \leq M(\gamma), \quad (i, j = 1, \dots, p).$$

The main result of this section is the following lemma.

LEMMA 8. *Given Assumptions 1–4, 6, and 7,*

$$p \lim_{n \rightarrow \infty} (1/n) R_n(\gamma_o) = -G_o(\gamma_o).$$

## 10. FINAL LEMMA

One more result is needed before we can prove the asymptotic normality of the ML estimator. To establish this result we shall make the following two additional assumptions.

Assumption 8. For every  $\gamma \in \overset{o}{\Gamma}$ , (i) there exists a finite positive number  $M(\gamma)$  such that for every  $n \in \mathbb{N}$ ,

$$(1/n) \text{tr}(\Omega_n(\gamma)) \leq M(\gamma),$$

and (ii)

$$\lim_{n \rightarrow \infty} (1/n^2) \text{tr}(\Omega_n^2(\gamma)) = 0.$$



Assumption 9. For every  $\phi \in \overset{o}{\Gamma}$  and  $\alpha > 0$ , there exists a neighborhood  $N(\phi) \subset \overset{o}{\Gamma}$  and an integer  $n_o$  (depending only on  $\alpha$  and  $N(\phi)$ ) such that the following ten conditions hold for every  $n > n_o$ ,  $\gamma \in N(\phi)$ , and  $i, j = 1, \dots, p$ :

- (i)  $(1/n) \left| \left( \mu_{(n)}(\gamma) - \mu_{(n)}(\phi) \right)' \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) \right| \leq \alpha,$
- (ii)  $(1/n) \left| \left( \mu_{(n)}(\gamma) - \mu_{(n)}(\phi) \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq \alpha,$
- (iii)  $(1/n) \left| \text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) - \text{tr} \left( \frac{\partial \Omega_n^{-1}(\phi)}{\partial \gamma_i} \Omega_n(\phi) \frac{\partial \Omega_n^{-1}(\phi)}{\partial \gamma_j} \Omega_n(\phi) \right) \right| \leq \alpha,$
- (iv)  $(1/n) \left| \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right) - \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\phi)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\phi) \right) \right| \leq \alpha,$
- (v)  $(1/n) \left| \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) - \left( \frac{\partial \mu_{(n)}(\phi)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\phi) \left( \frac{\partial \mu_{(n)}(\phi)}{\partial \gamma_j} \right) \right| \leq \alpha,$
- (vi)  $(1/n) \left| (\mu_{(n)}(\gamma) - \mu_{(n)}(\phi))' \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} (\mu_{(n)}(\gamma) - \mu_{(n)}(\phi)) \right| \leq \alpha,$
- (vii)  $(1/n) (\mu_{(n)}(\gamma) - \mu_{(n)}(\phi))' \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)^2 (\mu_{(n)}(\gamma) - \mu_{(n)}(\phi)) \leq \alpha,$
- (viii)  $(1/n) \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} - \frac{\partial \Omega_n^{-1}(\phi)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\phi)}{\partial \gamma_j} \right)' \times \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} - \frac{\partial \Omega_n^{-1}(\phi)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\phi)}{\partial \gamma_j} \right) \leq \alpha,$
- (ix)  $(1/n) \left( \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \Omega_n^{-1}(\phi) \frac{\partial^2 \mu_{(n)}(\phi)}{\partial \gamma_i \partial \gamma_j} \right)' \times \left( \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \Omega_n^{-1}(\phi) \frac{\partial^2 \mu_{(n)}(\phi)}{\partial \gamma_i \partial \gamma_j} \right) \leq \alpha,$
- (x)  $\max_{1 \leq i \leq n} \left| \lambda_i \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \Omega_n^{-1}(\phi)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq \alpha.$

At first glance the conditions in Assumption 9 may appear to be overly restrictive; in fact, they are not. We already know that for every  $\phi \in \overset{o}{\Gamma}$ ,  $\alpha > 0$ , and  $n \in \mathbb{N}$ , there exists a neighborhood  $N_n(\phi) \subset \overset{o}{\Gamma}$  such that all ten conditions of Assumption 9 hold for every  $\gamma \in N_n(\phi)$ . All that is required in addition is that the intersection  $\bigcap_{n=1}^{\infty} N_n(\phi)$  does not degenerate to a single



point  $\phi$ , but remains a true neighborhood (however small) of  $\phi$ . Notice that Assumption 9 is a local, not a global, condition: we require that it holds for points close to  $\phi$ , but not necessarily for every point in  $\Gamma$ .

Nevertheless, to verify Assumption 9 in practical situations, i.e., when  $\mu$  and  $\Omega$  are specified, can be arduous. Therefore we shall provide later (Theorem 3) conditions which are slightly stronger than Assumption 9 but easier to verify.

Let us now use Assumptions 8 and 9 to prove Lemma 9.

**LEMMA 9.** *Given Assumptions 1–3 and 7–9, there exists, for every  $\alpha > 0$ , a neighborhood  $N(\gamma_o) \subset \Gamma$  of  $\gamma_o$  such that for  $i, j = 1, \dots, p$ ,*

$$\lim_{n \rightarrow \infty} P \left[ (1/n) \sup_{\gamma \in N(\gamma_o)} |R_{nij}(\gamma) - R_{nij}(\gamma_o)| > \alpha \right] = 0.$$

## 11. THE CENTRAL LIMIT THEOREM FOR NORMALLY DISTRIBUTED (BUT DEPENDENT) OBSERVATIONS

We can now prove our main result.

**THEOREM 2.** *Suppose that Assumptions 1–9 are all satisfied. Suppose further that an ML estimator  $\{\hat{\gamma}_n(y_{(n)})\}$  exists asymptotically almost surely, and is weakly consistent. Then the sequence  $\{\hat{\gamma}_n(y_{(n)})\}$  is first-order efficient and asymptotically normally distributed, i.e.,*

$$\sqrt{n}(\hat{\gamma}_n(y_{(n)}) - \gamma_o) \xrightarrow{L} N(0, G_o^{-1}(\gamma_o)).$$

**Proof.** Lemmas 1–9 imply that conditions 1–9 of Theorem 1 are all satisfied. Thus the result follows. ■

## 12. A SET OF STRONGER ASSUMPTIONS

Some of the assumptions that underlie Theorem 2, notably Assumption 9, while weak and verifiable in principle, are still unappealing. For many purposes (for example in the case of first-order autocorrelation, see section 14) they are also unnecessarily general. Let us replace Assumptions 2 and 6–9 with the following set of assumptions which, while somewhat stronger, are more appealing and easier to verify.

**Assumption 2'.** For every (fixed)  $n \in \mathbb{N}$ ,  $\mu_{(n)}: \Gamma \rightarrow \mathbb{R}^n$  is a known three times differentiable vector function on  $\Gamma$  and  $\Omega_{(n)}: \Gamma \rightarrow \mathbb{R}^{n \times n}$  is a known twice continuously differentiable matrix function on  $\Gamma$ .



Assumption 6'. There exists a continuous function  $M: \overset{o}{\Gamma} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$  and  $\gamma \in \overset{o}{\Gamma}$ ,

- (i)  $(1/n) \operatorname{tr}(\Omega_n^2(\gamma)) \leq M(\gamma),$
- (ii)  $(1/n) \operatorname{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \right)^4 \leq M(\gamma), \quad (i = 1, \dots, p),$
- (iii)  $(1/n) \operatorname{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right)^2 \leq M(\gamma), \quad (i, j = 1, \dots, p).$

Assumption 7'. There exists a continuous function  $M: \overset{o}{\Gamma} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$  and  $\gamma \in \overset{o}{\Gamma}$ ,

- (i)  $(1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M(\gamma), \quad (i = 1, \dots, p),$
- (ii)  $(1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M(\gamma),$   
 $(i, j = 1, \dots, p),$
- (iii)  $(1/n) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \leq M(\gamma), \quad (i, j = 1, \dots, p),$
- (iv)  $(1/n) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)' \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_k} \right)^2 \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \leq M(\gamma),$   
 $(i, j, k = 1, \dots, p),$
- (v)  $(1/n) \left( \frac{\partial^3 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} \right)' \Omega_n^{-2}(\gamma) \left( \frac{\partial^3 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} \right) \leq M(\gamma), \quad (i, j, k = 1, \dots, p).$

Assumption 8'. There exist real-valued functions  $M_1$  and  $M_2$  defined and continuous on  $\overset{o}{\Gamma} \times \overset{o}{\Gamma}$  such that for all  $n \in \mathbb{N}$ ,  $\gamma \in \overset{o}{\Gamma}$  and  $\theta \in \overset{o}{\Gamma}$ ,

- (i)  $(1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\theta) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M_1(\gamma, \theta), \quad (i = 1, \dots, p),$
- (ii)  $(1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \left( \frac{\partial^2 \Omega_n^{-1}(\theta)}{\partial \gamma_j \partial \gamma_k} \right)^2 \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M_1(\gamma, \theta),$   
 $(i, j, k = 1, \dots, p),$
- (iii)  $\max_{1 \leq i \leq n} \left| \lambda_i \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \Omega_n^{-1}(\theta)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq M_2(\gamma, \theta), \quad (i, j = 1, \dots, p),$

with  $M_2(\gamma, \theta) \rightarrow 0$  as  $\gamma \rightarrow \theta$ .



Discussion. In Assumption 2' we require that the mean vector  $\mu$  is three times differentiable rather than twice continuously differentiable. The only reason for strengthening Assumption 2 in this way is that it enables us to replace Assumption 9 (ix) by the much simpler Assumption 7' (v), as we shall see in the proof of Theorem 3. If  $\mu$  is not three times differentiable, we can retain Assumption 9 (ix).

The eight conditions in Assumptions 6' and 7' are all of the following form: "there exists a real-valued function  $M$ , defined and continuous on  $\overset{o}{\Gamma}$ , such that  $(1/n)\eta_n^2(\gamma) \leq M(\gamma)$  for all  $n \in \mathbb{N}$  and  $\gamma \in \overset{o}{\Gamma}$ ". This implies that for every  $\gamma \in \overset{o}{\Gamma}$  a neighborhood  $N(\gamma)$  can be found such that the sequence of functions  $\{\eta_n^2\}$  is *uniformly bounded* on  $N(\gamma)$ . That is, it means that  $\{\eta_n^2\}$  is *locally uniformly bounded*. Similar remarks apply to Assumption 8'.

In the *linear* regression model  $y = X\gamma_o + \varepsilon$ ,  $\varepsilon \simeq N(0, \Omega(\gamma_o))$ , Assumptions 7'(iii)–(v) are trivially satisfied.

If there exists a  $\gamma \in \overset{o}{\Gamma}$  such that  $\Omega_n(\gamma) = I_n$  (which is almost always the case), then Assumption 7'(i) follows from Assumption 8'(i).

Let us now state Theorem 3.

**THEOREM 3.** *Suppose that Assumptions 1, 2', 3–5, 6'–8' are all satisfied. Suppose further that an ML estimator  $\{\hat{\gamma}_n(y_{(n)})\}$  exists asymptotically almost surely, and is weakly consistent. Then the sequence  $\{\hat{\gamma}_n(y_{(n)})\}$  is first-order efficient and asymptotically normally distributed, i.e.,*

$$\sqrt{n}(\hat{\gamma}_n(y_{(n)}) - \gamma_o) \xrightarrow{L} N(0, G_o^{-1}(\gamma_o)).$$

### 13. THE CASE OF UNIFORMLY BOUNDED EIGENVALUES

In the frequently encountered case where the eigenvalues of  $\Omega$ ,  $\Omega^{-1}$ , and its first and second derivatives are known to be locally uniformly bounded, the conditions simplify considerably. The following set of assumptions then replaces Assumptions 6'–8'.

Assumptions 6''. There exists a continuous function  $M: \overset{o}{\Gamma} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$  and  $\gamma \in \overset{o}{\Gamma}$ ,

$$(i) \quad \max_{1 \leq t \leq n} \lambda_t(\Omega_n(\gamma)) \leq M(\gamma),$$

$$(ii) \quad \max_{1 \leq t \leq n} \lambda_t(\Omega_n^{-1}(\gamma)) \leq M(\gamma),$$

$$(iii) \quad \max_{1 \leq t \leq n} \left| \lambda_t \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \right) \right| \leq M(\gamma), \quad (i = 1, \dots, p),$$

$$(iv) \quad \max_{1 \leq t \leq n} \left| \lambda_t \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq M(\gamma), \quad (i, j = 1, \dots, p).$$



Assumption 7''. There exists a continuous function  $M: \overset{o}{\Gamma} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$  and  $\gamma \in \overset{o}{\Gamma}$ ,

$$(i) \quad (1/n) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right) \leq M(\gamma), \quad (i = 1, \dots, p),$$

$$(ii) \quad (1/n) \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)' \left( \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right) \leq M(\gamma), \quad (i, j = 1, \dots, p),$$

$$(iii) \quad (1/n) \left( \frac{\partial^3 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} \right)' \left( \frac{\partial^3 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} \right) \leq M(\gamma), \quad (i, j, k = 1, \dots, p).$$

Assumption 8''. There exists a continuous function  $M_2: \overset{o}{\Gamma} \times \overset{o}{\Gamma} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$ ,  $\gamma \in \overset{o}{\Gamma}$ , and  $\theta \in \overset{o}{\Gamma}$ ,

$$\max_{1 \leq t \leq n} \left| \lambda_t \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \Omega_n^{-1}(\theta)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq M_2(\gamma, \theta), \quad (i, j = 1, \dots, p),$$

with  $M_2(\gamma, \theta) \rightarrow 0$  as  $\gamma \rightarrow \theta$ .

**THEOREM 4.** Suppose that Assumptions 1, 2', 3–5, 6''–8'' are all satisfied. Suppose further that an ML estimator  $\{\hat{\gamma}_n(y_{(n)})\}$  exists asymptotically almost surely, and is weakly consistent. Then the sequence  $\{\hat{\gamma}_n(y_{(n)})\}$  is first-order efficient and asymptotically normally distributed, i.e.,

$$\sqrt{n}(\hat{\gamma}_n(y_{(n)}) - \gamma_o) \xrightarrow{L} N(0, G_o^{-1}(\gamma_o)).$$

*Proof.* Immediate from Theorem 3, making repeated use of the fact that  $|x'Ax| \leq x'x \max_{1 \leq t \leq n} |\lambda_t(A)|$  for any symmetric matrix  $A$  and vector  $x$ . ■

## 14. AN EXAMPLE: FIRST-ORDER AUTOCORRELATION

By discussing a relatively simple example (first-order autocorrelation), we now hope to convince the reader that the conditions of Theorem 4 are easy to verify in practice, and lead to conditions which are weaker than those known from the literature (in the case of first-order autocorrelation, in particular Hildreth [8]). We wish to demonstrate Theorem 5.

**THEOREM 5.** Let  $\{y_1, y_2, \dots\}$  be a sequence of random variables, and assume that

C1. for every (fixed)  $n \in \mathbb{N}$ ,  $y_{(n)} := (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$  follows an  $n$ -variate normal distribution,

$$y_{(n)} \simeq N(\mu_{(n)}(\beta_o), \sigma_o^2 V_n(\rho_o)),$$



- where  $V_n(\rho)$  is symmetric  $n \times n$  matrix whose  $ij$ -th element is given by  $\rho^{|i-j|}/(1 - \rho^2)$ , and  $\beta_o \in \overset{\circ}{B} \subset \mathbb{R}^k$ ,  $\rho_o \in U := (-1, 1)$ , and  $\sigma_o^2 \in \mathbb{R}^+ := (0, \infty)$  are the true values of the  $k + 2$  parameters  $\beta := (\beta_1, \dots, \beta_k)$ ,  $\rho$  and  $\sigma^2$  to be estimated;
- C2. for every (fixed)  $n \in \mathbb{N}$ ,  $\mu_{(n)}: B \rightarrow \mathbb{R}^n$  is a known three times differentiable vector function on  $\overset{\circ}{B}$ , the interior of  $B$ ;
- C3. for every  $\beta \in \overset{\circ}{B}$  and  $\rho \in U$ , there exists a finite positive definite  $k \times k$  matrix  $H_o(\beta, \rho)$  such that

$$\lim_{n \rightarrow \infty} (1/n) \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta'} \right)' V_n^{-1}(\rho) \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta'} \right) = H_o(\beta, \rho);$$

- C4. there exists a continuous function  $M: \overset{\circ}{B} \rightarrow \mathbb{R}$  such that for all  $n \in \mathbb{N}$  and  $\beta \in \overset{\circ}{B}$ ,

$$\begin{aligned} \text{(i)} \quad & (1/n) \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta_i} \right)' \left( \frac{\partial \mu_{(n)}(\beta)}{\partial \beta_i} \right) \leq M(\beta), \quad (i = 1, \dots, k), \\ \text{(ii)} \quad & (1/n) \left( \frac{\partial^2 \mu_{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \right)' \left( \frac{\partial^2 \mu_{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \right) \leq M(\beta), \quad (i, j = 1, \dots, k), \\ \text{(iii)} \quad & (1/n) \left( \frac{\partial^3 \mu_{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \right)' \left( \frac{\partial^3 \mu_{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \right) \leq M(\beta), \quad (i, j, l = 1, \dots, k). \end{aligned}$$

Let  $\gamma := (\beta, \rho, \sigma^2)$  denote the vector of  $k + 2$  parameters and let  $\gamma_o := (\beta_o, \rho_o, \sigma_o^2)$ . Assume that an ML estimator  $\{\hat{\gamma}_n(y_{(n)})\}$  exists asymptotically almost surely, and is weakly consistent. Then the sequence  $\{\hat{\gamma}_n(y_{(n)})\}$  is first-order efficient and asymptotically normally distributed, i.e.,

$$\sqrt{n}(\hat{\gamma}_n(y_{(n)}) - \gamma_o) \xrightarrow{L} N(0, W(\gamma_o))$$

with

$$W(\gamma_o) := \begin{pmatrix} \sigma_o^2 H_o^{-1}(\beta_o, \rho_o) & 0 & 0 \\ 0' & (1 - \rho_o^2) & 0 \\ 0' & 0 & 2\sigma_o^4 \end{pmatrix}.$$

**Discussion.** We have assumed that the structural parameters  $\beta$  and the covariance parameters  $\sigma^2$  and  $\rho$  are functionally unrelated. This restriction can easily be removed.

Let  $\eta_n^i(\beta) := (1/n)(\partial \mu_{(n)}(\beta)/\partial \beta_i)'(\partial \mu_{(n)}(\beta)/\partial \beta_i)$ . From C3 we obtain  $\eta_n^i(\beta) \rightarrow H_o^{ii}(\beta, 0) > 0$  as  $n \rightarrow \infty$ , for every  $\beta \in \overset{\circ}{B}$ . ( $H_o^{ii}$  denotes the  $i$ -th diagonal element of  $H_o$ .) This implies that, for every  $\beta \in \overset{\circ}{B}$ , the sequence  $\{\eta_n^i(\beta)\}$  is bounded. It does not imply that for every  $\beta \in \overset{\circ}{B}$  a neighborhood  $N(\beta)$  can be found such that the sequence of functions  $\{\eta_n^i\}$  is uniformly bounded on  $N(\beta)$ . This last requirement, however, is what we need, and condition C4(i) guarantees the existence of such a neighborhood. Notice again that C4(i) implies that  $\{\eta_n^i\}$



is *locally* uniformly bounded; we do not require that it is globally uniformly bounded. Similar remarks apply to conditions C4(ii) and (iii).

In the special case of the *linear* regression model

$$y = X\beta_o + \varepsilon, \quad \varepsilon \simeq N(0, \sigma_o^2 V(\rho_o)), \quad (28)$$

conditions C2 and C4 are redundant, and condition C3 boils down to

C3'. for every  $\rho \in U$ , the matrix  $(1/n)X'V^{-1}(\rho)X$  converges, as  $n \rightarrow \infty$ , to a positive definite  $k \times k$  matrix  $H_o(\rho)$ .

Let us compare our results with those of Hildreth [8]. Hildreth considers the linear model of equation (28). His conditions for asymptotic normality (p. 584) consist of our conditions C1 and C3', and two additional ones involving the matrix  $X = (x_{ti})$ , namely (i)  $|x_{ti}|$  is bounded ( $i = 1, \dots, k$ ;  $t = 1, 2, \dots$ ), and (ii)  $(1/n)\sum_{t=\tau}^n x_{ti}x_{t-\tau+1,j}$  converges, as  $n \rightarrow \infty$ , to a finite limit ( $i, j = 1, \dots, k$ ;  $\tau = 1, \dots, n$ ). We conclude that our conditions are substantially weaker than those of Hildreth.

## NOTES

1. "Everybody believes in the law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact." Poincaré [15, p. 171] attributes this remark to Lippmann. Also quoted by Cramér [3, p. 232].

2. Since we have adopted the convention that each expectation is evaluated under the true value  $\gamma_o$ , it would be incorrect to define  $G_n(\gamma) := -ER_n(\gamma)$ .

3. See also Holly [9].

4. We use the fact that  $\text{tr} ABCB = (\text{vec } A)'(B \otimes B) \text{vec } C$  for symmetric matrices  $A$ ,  $B$ , and  $C$  of the same order.

5. See also Magnus [10, Lemma 5] and Holly [9].

6. A square matrix is said to be unit upper triangular if its diagonal elements are 1 and all elements below the diagonal are zero.

7. Lemma A.4 replaces an earlier and less general one. We are indebted to Peter Robinson for suggesting this lemma and providing the proof.

8. See also formula (37) in Magnus [10, p. 306] and the correction on p. 261.

## REFERENCES

1. Anderson, T. W. *The statistical analysis of time series*. New York: John Wiley, 1971.
2. Apostol, T. M. *Mathematical analysis* (2nd edition). Reading, Mass.: Addison-Wesley, 1974.
3. Cramér, H. *Mathematical methods of statistics*. Princeton: Princeton University Press, 1946.
4. Hannan, E. J. *Multiple time series*. New York: John Wiley, 1970.
5. Hannan, E. J. The asymptotic theory of linear time-series models. *Journal of Applied Probability* 10 (1973): 130-145.
6. Heijmans, R.D.H. and J. R. Magnus. On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Statistica Neerlandica* 40 (1986): 169-188.
7. Heijmans, R.D.H. and J. R. Magnus. Consistent maximum-likelihood estimation with dependent observations: the general (non-normal) case and the normal case. *Journal of Econometrics* 32 (1986): 253-285.



8. Hildreth, C. Asymptotic distribution of maximum likelihood estimators in a linear model with autoregressive disturbances. *Annals of Mathematical Statistics* 40 (1969): 583–594.
9. Holly, A. Problem 85.1.2. *Econometric Theory* 1 (1985): 143–144. Solution in *Econometric Theory* 2 (1986): forthcoming.
10. Magnus, J. R. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics* 7 (1978): 281–312. Corrigenda in: *Journal of Econometrics* 10 (1979): 261.
11. Magnus, J. R. and H. Neudecker. The commutation matrix: some properties and applications. *The Annals of Statistics* 7 (1979): 381–394.
12. Malinvaud, E. *Statistical methods of econometrics* (2nd edition). Amsterdam: North-Holland, 1970.
13. Mann, H. B. and A. Wald. On the statistical treatment of linear stochastic difference equations. *Econometrica* 11 (1943): 173–220.
14. Marcus, M. and H. Minc. *A survey of matrix theory and matrix inequalities*. Boston: Allyn and Bacon, 1964.
15. Poincaré, H. *Calcul des probabilités* (2nd edition). Paris, 1912.
16. Rao, C. R. *Linear statistical inference and its applications* (2nd edition). New York: John Wiley, 1973.
17. Rozanov, Yu. A. *Stationary random processes*. San Francisco: Holden-Day, 1967.
18. Rubin, H. Some results on the asymptotic distribution of maximum- and quasi-maximum-likelihood estimates (abstract). *Annals of Mathematical Statistics* 19 (1948): 598.
19. Schönfeld, P. A useful central limit theorem for  $m$ -dependent variables. *Metrika* 17 (1971): 116–128.
20. Weiss, L. Asymptotic properties of maximum likelihood estimators in some nonstandard cases, II. *Journal of the American Statistical Association* 68 (1973): 428–430.



# APPENDIX A: SEVEN AUXILIARY LEMMAS

In this Appendix we prove seven lemmas which are applied in the main text and in Appendix B.

LEMMA A.1. (Anderson [1, p. 25]). *Let  $\mu_1, \mu_2, \dots$  be a sequence of nonnegative real numbers. Then  $(1/n) \max_{1 \leq t \leq n} \mu_t \rightarrow 0$  as  $n \rightarrow \infty$  if and only if  $(1/n)\mu_n \rightarrow 0$  as  $n \rightarrow \infty$ .*

Proof. If  $(1/n) \max_{1 \leq t \leq n} \mu_t \rightarrow 0$ , then  $(1/n)\mu_n \leq (1/n) \max_{1 \leq t \leq n} \mu_t \rightarrow 0$ . To prove the converse, assume that  $(1/n)\mu_n \rightarrow 0$ . Let  $t(n)$  be the largest index  $t$  ( $1 \leq t \leq n$ ) such that

$$\mu_{t(n)}$$

as  $n \rightarrow \infty$ , then  $(1/n) \max_{1 \leq t \leq n} \mu_t \leq (1/t(n))\mu_{t(n)} \rightarrow 0$ . ■

LEMMA A.2. *Let  $\mu_1, \mu_2, \dots$  be a sequence of nonnegative real numbers. Suppose that  $(1/n) \sum_{i=1}^n \mu_i \rightarrow \mu_0$  (finite), as  $n \rightarrow \infty$ . Then*

$$(1/n) \max_{1 \leq t \leq n} \mu_t \rightarrow 0, \quad \text{as } n \rightarrow \infty$$

Proof. Immediate from Lemma A.1. ■

LEMMA A.3. *Let  $\{x_n, n \in \mathbb{N}\}$  be a sequence of random variables with finite second moments. If there exists a sequence of random variables  $\{y_n, n \in \mathbb{N}\}$  such that*

$$x_n = x_n(y_1, \dots, y_n)$$

and

$$E(x_n | y_1, \dots, y_{n-1}) = 0,$$

then the random variables  $\{x_n, n \in \mathbb{N}\}$  are uncorrelated.

Proof. For  $i < j$ ,  $\text{cov}(x_i, x_j) = E x_i x_j = E E(x_i x_j | y_1, \dots, y_{j-1}) = E [x_i E(x_j | y_1, \dots, y_{j-1})] = 0$ . ■

LEMMA A.4.<sup>7</sup> *Let  $v_1, v_2, \dots$  be a sequence of random variables satisfying  $E|v_i|^\delta \leq C < \infty$  for some  $\delta > 0$ , and let  $\mu_1, \mu_2, \dots$  be a sequence of nonnegative real numbers. If there exists an  $\alpha > 0$  such that*

$$n^{-\alpha} \mu_n \rightarrow 0 \quad \text{as } n \rightarrow \infty$$



and

$n^{-\alpha\theta} \sum_{t=1}^n \mu_t^\theta$  is bounded in  $n$  for some  $\theta \in (0, \delta)$ ,

then

$$\lim_{n \rightarrow \infty} E \left| n^{-\alpha} \max_{1 \leq t \leq n} \mu_t v_t \right|^\theta = 0.$$

Proof. Let  $d_n := n^{-\alpha} \max_{1 \leq t \leq n} \mu_t$ . Let  $\varepsilon > 0$  be arbitrary, and define

$$v'_t := \begin{cases} v_t, & \text{if } |v_t| \geq \varepsilon/d_n, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} E \left( \max_{1 \leq t \leq n} \mu_t^\theta |v'_t|^\theta \right) &\leq \sum_{t=1}^n \mu_t^\theta E |v'_t|^\theta = \sum_{t=1}^n \mu_t^\theta E (|v'_t|^{-(\delta-\theta)} |v'_t|^\delta) \\ &\leq \sum_{t=1}^n \mu_t^\theta (d_n/\varepsilon)^{\delta-\theta} E |v_t|^\delta \leq C(d_n/\varepsilon)^{\delta-\theta} \sum_{t=1}^n \mu_t^\theta. \end{aligned}$$

Hence,

$$\begin{aligned} E \left| n^{-\alpha} \max_{1 \leq t \leq n} \mu_t v_t \right|^\theta &\leq n^{-\alpha\theta} E \left( \max_{1 \leq t \leq n} \mu_t |v_t| \right)^\theta \\ &\leq n^{-\alpha\theta} E \left( \max_{1 \leq t \leq n} \mu_t^\theta |v'_t|^\theta \right) + n^{-\alpha\theta} E \left( \max_{1 \leq t \leq n} \mu_t^\theta |v_t - v'_t|^\theta \right) \\ &\leq C(d_n/\varepsilon)^{\delta-\theta} n^{-\alpha\theta} \sum_{t=1}^n \mu_t^\theta + n^{-\alpha\theta} \max_{1 \leq t \leq n} \mu_t^\theta (\varepsilon/d_n)^\theta \\ &\leq C'(d_n/\varepsilon)^{\delta-\theta} + \varepsilon^\theta. \end{aligned}$$

Since  $d_n \rightarrow 0$  as  $n \rightarrow \infty$  (using Lemma A.1), and  $\varepsilon$  was chosen arbitrarily, the result follows. ■

**LEMMA A.5.** Let  $v_1, v_2, \dots$  be a sequence of  $\chi^2(1)$  distributed random variables, not necessarily independent, and let  $\mu_1, \mu_2, \dots$  be a sequence of nonnegative real numbers. Suppose that  $(1/n) \sum_{t=1}^n \mu_t \rightarrow \mu_0$  (finite), as  $n \rightarrow \infty$ . Then

$$p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \mu_t v_t^\alpha = 0$$

for every  $\alpha > 0$ .



Proof. Immediate from Lemma A.4. ■

LEMMA A.6. Let  $A_n$  and  $\Omega_n$  be given  $n \times n$  matrices, positive definite for every  $n \in \mathbb{N}$ , and let  $\varepsilon_{(n)} \simeq N(0, \Omega_n)$ . Let  $\Gamma$  be an open set in  $\mathbb{R}^p$ , and let  $\gamma_o$  be a given point in  $\Gamma$ . For every (fixed)  $n \in \mathbb{N}$ , let  $\beta_n$  be a real-valued function,  $b_{(n)}$  an  $n \times 1$  vector function, and  $B_n$  an  $n \times n$  symmetric matrix function, each defined on  $\Gamma$ . Assume that for every  $\alpha > 0$  there exists a neighborhood  $N(\gamma_o) \subset \Gamma$  of  $\gamma_o$  and an integer  $n_o$  (depending only on  $\alpha$  and  $N(\gamma_o)$ ) such that for every  $n > n_o$  and  $\gamma \in N(\gamma_o)$ ,

- (i)  $(1/n)|\beta_n(\gamma)| \leq \alpha$ ,
- (ii)  $(1/n)b'_{(n)}(\gamma)A_nb_{(n)}(\gamma) \leq \alpha$ ,
- (iii)  $\max_{1 \leq t \leq n} |\lambda_{nt}(\gamma)| \leq \alpha$ ,

where  $\lambda_{nt}(\gamma)$ ,  $t = 1, \dots, n$ , denote the eigenvalues of the symmetric  $n \times n$  matrix  $A_n^{1/2}B_n(\gamma)A_n^{1/2}$

Assume further that

- (iv)  $(1/n) \operatorname{tr}(A_n^{-1}\Omega_n) \leq K$ ,  $(n \in \mathbb{N})$ ,

for some finite positive number  $K$ , and

- (v)  $\lim_{n \rightarrow \infty} (1/n^2) \operatorname{tr}(A_n^{-1}\Omega_n)^2 = 0$ .

Then there exists for every  $\alpha > 0$  a neighborhood  $N^*(\gamma_o) \subset \Gamma$  of  $\gamma_o$  such that

$$\lim_{n \rightarrow \infty} P \left[ (1/n) \sup_{\gamma \in N^*(\gamma_o)} \left| \beta_n(\gamma) + b'_{(n)}(\gamma)\varepsilon_{(n)} + \varepsilon'_{(n)}B_n(\gamma)\varepsilon_{(n)} \right| > \alpha \right] = 0.$$

Note. In practice we shall choose either  $A_n = I_n$  or  $A_n = \Omega_n$ . In the latter case conditions (iv) and (v) are automatically satisfied for  $K = 1$ .

Proof. Let  $\alpha > 0$  be arbitrary. Given conditions (i)–(iii), there exists a neighborhood  $N_o(\gamma_o) \subset \Gamma$  of  $\gamma_o$  and an integer  $n_o$  (depending on  $\alpha$  and  $N_o(\gamma_o)$ ) such that, for every  $n > n_o$  and  $\gamma \in N_o(\gamma_o)$ ,

$$\begin{aligned} (1/n)|\beta_n(\gamma)| &\leq \alpha/2, \\ (1/n)b'_{(n)}(\gamma)A_nb_{(n)}(\gamma) &\leq \alpha^2/(32K), \end{aligned}$$

and

$$\max_{1 \leq t \leq n} |\lambda_{nt}(\gamma)| \leq \alpha/(8K).$$



Let  $N_1(\gamma_o)$  be another neighborhood of  $\gamma_o$  such that its closure  $\bar{N}_1(\gamma_o)$  is a subset of  $N_o(\gamma_o)$ . Then

$$(1/n) \sup_{\gamma \in N_1(\gamma_o)} |\beta_n(\gamma)| \leq \alpha/2, \quad (n > n_o), \quad (\text{A.1})$$

$$(1/n) \sup_{\gamma \in N_1(\gamma_o)} b'_{(n)}(\gamma) A_n b_{(n)}(\gamma) \leq \alpha^2/(32K), \quad (n > n_o), \quad (\text{A.2})$$

and

$$\sup_{\gamma \in N_1(\gamma_o)} \max_{1 \leq t \leq n} |\lambda_{nt}(\gamma)| \leq \alpha/(8K), \quad (n > n_o). \quad (\text{A.3})$$

To avoid cumbersome notation let us abbreviate  $\varepsilon_{(n)}$ ,  $\beta_n(\gamma)$ ,  $b_{(n)}(\gamma)$ , and  $B_n(\gamma)$  to  $\varepsilon$ ,  $\beta$ ,  $b$ , and  $B$ ; also, each supremum is understood to be over all  $\gamma$  in  $N_1(\gamma_o)$ . Now consider the random variable  $w_n := (1/n)\varepsilon' A^{-1}\varepsilon$ , whose distribution does *not* depend on  $\gamma$ , and note from (iv) and (v) that

$$Ew_n \leq K \quad \text{and} \quad \text{var } w_n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Let  $\phi_n := (1/n) \sup b' A b$ . Then, since

$$(b'\varepsilon)^2 \leq (b' A b)(\varepsilon' A^{-1}\varepsilon),$$

we obtain

$$\begin{aligned} P((1/n) \sup |b'\varepsilon| > \alpha/4) &\leq P(\sqrt{\phi_n w_n} > \alpha/4) \\ &= P(w_n > \alpha^2/(16\phi_n)) \leq P(w_n > 2K), \end{aligned}$$

using (A.2). Similarly, let  $\psi_n := \sup \max_{1 \leq t \leq n} |\lambda_{nt}(\gamma)|$ . Then, since

$$|\varepsilon' B \varepsilon| \leq \max_{1 \leq t \leq n} |\lambda_{nt}(\gamma)| (\varepsilon' A^{-1}\varepsilon),$$

we obtain

$$\begin{aligned} P((1/n) \sup |\varepsilon' B \varepsilon| > \alpha/4) &\leq P(\psi_n w_n > \alpha/4) \\ &= P(w_n > \alpha/(4\psi_n)) \leq P(w_n > 2K), \end{aligned}$$

using (A.3). Hence, for  $n > n_o$ ,

$$\begin{aligned} P((1/n) \sup |\beta + b'\varepsilon + \varepsilon' B \varepsilon| > \alpha) &\leq P((1/n) \sup |\beta| + (1/n) \sup |b'\varepsilon| + (1/n) \sup |\varepsilon' B \varepsilon| > \alpha) \\ &\leq P((1/n) \sup |b'\varepsilon| + (1/n) \sup |\varepsilon' B \varepsilon| > \alpha/2) \quad (\text{using A.1}) \\ &\leq P(\max \{(1/n) \sup |b'\varepsilon|, (1/n) \sup |\varepsilon' B \varepsilon|\} > \alpha/4) \\ &\leq P((1/n) \sup |b'\varepsilon| > \alpha/4) + P((1/n) \sup |\varepsilon' B \varepsilon| > \alpha/4) \\ &\leq 2P(w_n > 2K) \leq 2P(w_n - Ew_n > K) \\ &\leq 2P(|w_n - Ew_n| > K) \leq (2/K^2) \text{var } w_n \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . This concludes the proof. ■



LEMMA A.7. *Let  $u \simeq N(0, I_n)$ ,  $x := Au$ , and  $y := Bu$ , where  $A$  and  $B$  are real  $m \times n$  matrices. Then*

$$\text{var } x'y \leq \frac{1}{2}(\text{var } x'x + \text{var } y'y).$$

Proof. We have

$$\begin{aligned} \text{var}(x'x + y'y) - 4 \text{var } x'y &= \text{var } u'(A'A + B'B)u - \text{var } u'(A'B + B'A)u \\ &= 2 \text{tr}(A'A + B'B)^2 - 2 \text{tr}(A'B + B'A)^2 \\ &= 2 \text{tr}(A + B)'(A + B)(A - B)'(A - B) \geq 0, \end{aligned}$$

and hence

$$\begin{aligned} 2(\text{var } x'x + \text{var } y'y) &= \text{var}(x'x + y'y) + \text{var}(x'x - y'y) \\ &\geq \text{var}(x'x + y'y) \geq 4 \text{var } x'y. \end{aligned}$$

■



## APPENDIX B: PROOFS

Proof of Proposition 1. The loglikelihood  $\Lambda_n(\gamma)$  is given by equation (5). Since

$$\frac{\partial}{\partial \gamma_j} \log |\Omega_n(\gamma)| = -\text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right),$$

and

$$\begin{aligned} & \frac{\partial}{\partial \gamma_j} (y_{(n)} - \mu_{(n)}(\gamma))' A_n(\gamma) (y_{(n)} - \mu_{(n)}(\gamma)) \\ &= -2(y_{(n)} - \mu_{(n)}(\gamma))' A_n(\gamma) \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \\ & \quad + (y_{(n)} - \mu_{(n)}(\gamma))' \left( \frac{\partial A_n(\gamma)}{\partial \gamma_j} \right) (y_{(n)} - \mu_{(n)}(\gamma)), \end{aligned} \tag{B.1}$$

for any symmetric  $n \times n$  matrix function  $A$  defined and differentiable on  $\overset{o}{\Gamma}$ , we obtain the expression for  $l_{nj}(\gamma)$ . Further, since

$$\begin{aligned} & \frac{\partial}{\partial \gamma_i} \text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) = -\text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) \\ & \quad + \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right), \\ & \frac{\partial}{\partial \gamma_i} \left( (y_{(n)} - \mu_{(n)}(\gamma))' \Omega_n^{-1}(\gamma) \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) \\ &= -\left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) \\ & \quad + (y_{(n)} - \mu_{(n)}(\gamma))' \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} + \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right), \end{aligned}$$

and using (B.1) we obtain the expression for  $R_{nij}(\gamma)$ . Finally, since

$$E(y_{(n)} - \mu_{(n)}(\gamma_o)) = 0, \quad E(y_{(n)} - \mu_{(n)}(\gamma_o))(y_{(n)} - \mu_{(n)}(\gamma_o))' = \Omega_n(\gamma_o),$$

we obtain the expression for  $G_{nij}(\gamma_o)$ . ■

Proof of Proposition 2. Recall from Proposition 1 that

$$l_{nj}(\gamma_o) = \frac{1}{2} \text{tr} \left( \frac{\partial \Omega^{-1}(\gamma_o)}{\partial \gamma_j} \Omega(\gamma_o) \right) + \left( \frac{\partial \mu(\gamma_o)}{\partial \gamma_j} \right)' \Omega^{-1}(\gamma_o) \varepsilon - \frac{1}{2} \varepsilon' \frac{\partial \Omega^{-1}(\gamma_o)}{\partial \gamma_j} \varepsilon,$$



where  $\varepsilon := y - \mu(\gamma_o) \simeq N(0, \Omega(\gamma_o))$ . Clearly we have  $El_{nj}(\gamma_o) = 0$  (first-order regularity). To prove second-order regularity, we first note that

$$E(\varepsilon' A \varepsilon) \varepsilon = 0$$

and

$$\text{cov}(\varepsilon' A \varepsilon, \varepsilon' B \varepsilon) = 2 \text{tr}(A \Omega(\gamma_o) B \Omega(\gamma_o))$$

for any pair of symmetric  $n \times n$  matrices  $A$  and  $B$  (see, e.g., Magnus and Neudecker [11, Corollary 4.1]). It is then easy to show that

$$\begin{aligned} El_{ni}(\gamma_o) l_{nj}(\gamma_o) &= \text{cov}(l_{ni}(\gamma_o), l_{nj}(\gamma_o)) \\ &= \left( \frac{\partial \mu(\gamma_o)}{\partial \gamma_i} \right)' \Omega^{-1}(\gamma_o) \left( \frac{\partial \mu(\gamma_o)}{\partial \gamma_j} \right) \\ &\quad + \frac{1}{2} \text{tr} \left( \frac{\partial \Omega^{-1}(\gamma_o)}{\partial \gamma_i} \Omega(\gamma_o) \frac{\partial \Omega^{-1}(\gamma_o)}{\partial \gamma_j} \Omega(\gamma_o) \right) \\ &= -ER_{nij}(\gamma_o), \end{aligned}$$

according to Proposition 1. ■

**Proof of Lemma 4.** From equation (18) we have

$$\xi_{ij} = -\frac{1}{2} \alpha_{ij} (1 - u_t^2) + (\delta_{ij} - w_{ij}) u_t$$

and hence

$$\xi_{ij}^2 \leq \alpha_{ij}^2 + \alpha_{ij}^2 u_t^4 + 4 \delta_{ij}^2 u_t^2 + 4 w_{ij}^2 u_t^2.$$

Let

$$\tau_{ij}^2 := (1/\alpha_{ij}(\gamma_o)) (\partial z_t(\gamma_o) / \partial \gamma_j)' \Omega_t(\gamma_o) (\partial z_t(\gamma_o) / \partial \gamma_j). \tag{B.2}$$

Then the random variable  $w_{ij}$ , defined in equation (22), is distributed as  $N(0, \tau_{ij}^2)$ , and so the random variable  $v_{ij} := w_{ij}^2 / \tau_{ij}^2$  is distributed as  $\chi^2(1)$ . Thus we obtain

$$\xi_{ij}^2 \leq 4 \sigma_{ij}^2 (1 + u_t^2 + u_t^4 + v_{ij}^2),$$

where we recall from equation (26) that

$$\sigma_{ij}^2 := \text{var } \xi_{ij} = \frac{1}{2} \alpha_{ij}^2 + \delta_{ij}^2 + \tau_{ij}^2,$$

using (B.2). Since  $\{\xi_{ij}, 1 \leq i \leq n\}$  is a sequence of uncorrelated random variables (Lemma A.3 and equation (24)), we have

$$\sum_{i=1}^n \sigma_{ij}^2 = \sum_{i=1}^n \text{var } \xi_{ij} = \text{var } \sum_{i=1}^n \xi_{ij} = \text{var } l_{nj}(\gamma_o) = G_{njj}(\gamma_o),$$



using Propositions 1 and 2, so that

$$(1/n) \sum_{t=1}^n \sigma_{tj}^2 = (1/n) G_{njj}(\gamma_o) \rightarrow G_{ojj}(\gamma_o), \quad (\text{B.3})$$

as  $n \rightarrow \infty$  (Assumption 4). Hence

$$p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \xi_{tj}^2 \leq 4p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \sigma_{tj}^2 (1 + u_t^2 + u_t^4 + v_{tj}^2) = 0$$

using Lemmas A.2 and A.5 in Appendix A. ■

**Proof of Lemma 5.** Recall from equation (25) that

$$E(\xi_{ti} \xi_{tj} | \varepsilon_{(t-1)}) = \frac{1}{2} \alpha_{ti} \alpha_{tj} + (\delta_{ti} - w_{ti})(\delta_{tj} - w_{tj}).$$

Let us define the strictly upper triangular  $n \times n$  matrix

$$Q_{nj} := (\partial Z_n(\gamma_o) / \partial \gamma_j) A_n^{-1/2}(\gamma_o), \quad (j = 1, \dots, p),$$

and the  $n \times 1$  vector

$$p_{(n)j} := A_n^{-1/2}(\gamma_o) Z'_n(\gamma_o) (\partial \mu_n(\gamma_o) / \partial \gamma_j), \quad (j = 1, \dots, p),$$

whose components are  $\delta_{tj}$  ( $1 \leq t \leq n$ ). Then

$$\begin{aligned} \text{var} \sum_{t=1}^n E(\xi_{ti} \xi_{tj} | \varepsilon_{(t-1)}) &= \text{var} \sum_{t=1}^n (w_{ti} w_{tj} - (\delta_{ti} w_{tj} + \delta_{tj} w_{ti})) \\ &= \text{var} (\varepsilon'_{(n)} Q_{ni} Q'_{nj} \varepsilon_{(n)} - (p'_{(n)i} Q'_{nj} + p'_{(n)j} Q'_{ni}) \varepsilon_{(n)}) \\ &= \text{var} (\varepsilon'_{(n)} Q_{ni} Q'_{nj} \varepsilon_{(n)}) + \text{var} (p'_{(n)i} Q'_{nj} \varepsilon_{(n)} + p'_{(n)j} Q'_{ni} \varepsilon_{(n)}) \\ &\leq \frac{1}{2} \text{var} (\varepsilon'_{(n)} Q_{ni} Q'_{ni} \varepsilon_{(n)}) + \frac{1}{2} \text{var} (\varepsilon'_{(n)} Q_{nj} Q'_{nj} \varepsilon_{(n)}) \\ &\quad + 2 \text{var} (p'_{(n)i} Q'_{nj} \varepsilon_{(n)}) + 2 \text{var} (p'_{(n)j} Q'_{ni} \varepsilon_{(n)}) \end{aligned}$$

(using Lemma A.7)

$$\begin{aligned} &= \text{tr} (Q'_{ni} \Omega_n(\gamma_o) Q_{ni})^2 + \text{tr} (Q'_{nj} \Omega_n(\gamma_o) Q_{nj})^2 \\ &\quad + 2 p'_{(n)i} Q'_{nj} \Omega_n(\gamma_o) Q_{nj} p_{(n)i} + 2 p'_{(n)j} Q'_{ni} \Omega_n(\gamma_o) Q_{ni} p_{(n)j}. \end{aligned}$$

Now, since

$$\begin{aligned} p'_{(n)i} Q'_{nj} \Omega_n(\gamma_o) Q_{nj} p_{(n)i} &\leq (p'_{(n)i} p_{(n)i}) \max_{1 \leq t \leq n} \lambda_t(Q'_{nj} \Omega_n(\gamma_o) Q_{nj}) \\ &\leq G_{nii}(\gamma_o) \sqrt{\text{tr} (Q'_{nj} \Omega_n(\gamma_o) Q_{nj})^2}, \end{aligned}$$



we obtain

$$\begin{aligned} (1/n^2) \operatorname{var} \sum_{t=1}^n E(\xi_{ti}\xi_{tj} | \varepsilon_{(t-1)}) \\ \leq (1/n^2) \operatorname{tr}(Q'_{ni}\Omega_n(\gamma_o)Q_{ni})^2 + (1/n^2) \operatorname{tr}(Q'_{nj}\Omega_n(\gamma_o)Q_{nj})^2 \\ + (2/n)G_{nii}(\gamma_o)\sqrt{(1/n^2) \operatorname{tr}(Q'_{nj}\Omega_n(\gamma_o)Q_{nj})^2} \\ + (2/n)G_{njj}(\gamma_o)\sqrt{(1/n^2) \operatorname{tr}(Q'_{ni}\Omega_n(\gamma_o)Q_{ni})^2} \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ , by Assumptions 4 and 5. ■

Proof of Lemma 6. Let

$$v_{tij} := \xi_{ti}\xi_{tj} - E(\xi_{ti}\xi_{tj} | y_1, \dots, y_{t-1}), \quad (i, j = 1, \dots, p),$$

and recall from equation (27) that its variance is bounded by

$$\operatorname{var} v_{tij} \leq 14\sigma_{ii}^2\sigma_{ij}^2.$$

Then

$$\begin{aligned} (1/n^2) \sum_{t=1}^n \operatorname{var} v_{tij} &\leq (14/n^2) \sum_{t=1}^n \sigma_{ii}^2\sigma_{ij}^2 \\ &\leq 14((1/n) \max_{1 \leq i \leq n} \sigma_{ii}^2)((1/n) \sum_{t=1}^n \sigma_{ij}^2) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ , because of (B.3) and Lemma A.2. ■

Proof of Proposition 4. Let

$$\xi_n := \begin{bmatrix} \xi_{n1} \\ \vdots \\ \xi_{np} \end{bmatrix} = l_n(\gamma_o) - l_{n-1}(\gamma_o),$$

where  $l_o(\gamma_o) = 0$  since  $L_o(\gamma) = 1$  for all  $\gamma$ . We wish to demonstrate

$$n^{-1/2} \sum_{t=1}^n \xi_t \xrightarrow{L} N(0, G_o(\gamma_o)).$$

Now,  $\{l_n(\gamma_o)\}$  is a vector martingale (Proposition 3), so that  $\{n^{-1/2}\xi_t, 1 \leq t \leq n, n \in \mathbb{N}\}$  is a vector martingale difference array. (See Heijmans and Magnus [6, section 3] for details.) Hence, if we can show that

$$(i) \quad (1/n)E \max_{1 \leq i \leq n} \xi_{ij}^2 \text{ is bounded in } n, \quad (j = 1, \dots, p),$$



$$(ii) \quad p \lim_{n \rightarrow \infty} (1/n) \max_{1 \leq t \leq n} \xi_{tj}^2 = 0, \quad (j = 1, \dots, p),$$

$$(iii) \quad p \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n \xi_{ti} \xi_{tj} = G_{oij}(\gamma_o), \quad (i, j = 1, \dots, p),$$

then the result will follow from Proposition 1 of [6].

To prove (i), we note that

$$\begin{aligned} (1/n)E \max_{1 \leq t \leq n} \xi_{tj}^2 &\leq (1/n)E \sum_{t=1}^n \xi_{tj}^2 \\ &= (1/n) \sum_{t=1}^n \text{var } \xi_{tj} = (1/n)G_{njj}(\gamma_o) \rightarrow G_{ojj}(\gamma_o), \end{aligned}$$

as  $n \rightarrow \infty$ , using (B.3). Hence (i) holds. Condition (ii) follows from Lemma 4. Since  $E\xi_{ti}\xi_{sj} = 0$  ( $t \neq s$ ), we obtain

$$\begin{aligned} (1/n)E \sum_{t=1}^n \xi_{ti} \xi_{tj} &= (1/n)E \left( \sum_{t=1}^n \xi_{ti} \right) \left( \sum_{t=1}^n \xi_{tj} \right) \\ &= (1/n)El_{ni}(\gamma_o)l_{nj}(\gamma_o) = (1/n)G_{nij}(\gamma_o) \rightarrow G_{oij}(\gamma_o), \end{aligned}$$

as  $n \rightarrow \infty$ . Also, in view of Lemmas 5 and 6,

$$(1/n^2) \text{var} \sum_{t=1}^n \xi_{ti} \xi_{tj} \rightarrow 0,$$

as  $n \rightarrow \infty$ . Hence condition (iii) holds as well. ■

**Proof of Lemma 8.** We recall, from Proposition 1, that

$$\begin{aligned} R_{nij}(\gamma_o) &= -G_{nij}(\gamma_o) + q'_{nij}(\gamma_o)\varepsilon_{(n)} \\ &\quad - \frac{1}{2}(\varepsilon'_{(n)}B_{nij}(\gamma_o)\varepsilon_{(n)} - E\varepsilon'_{(n)}B_{nij}(\gamma_o)\varepsilon_{(n)}), \end{aligned}$$

where

$$q_{nij}(\gamma) := \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} + \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} + \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j},$$

and

$$B_{nij}(\gamma) := \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j}.$$

In view of Assumption 4, it is sufficient to show that

$$p \lim_{n \rightarrow \infty} (1/n)q'_{nij}\varepsilon_{(n)} = 0$$



and

$$p \lim_{n \rightarrow \infty} (1/n)(\varepsilon'_{(n)} B_{nij}(\gamma_o) \varepsilon_{(n)} - E \varepsilon'_{(n)} B_{nij}(\gamma_o) \varepsilon_{(n)}) = 0.$$

Sufficient (and in the first case also necessary) for these two conditions is that

$$(1/n^2) \text{var } q'_{nij} \varepsilon_{(n)} \rightarrow 0$$

and

$$(1/n^2) \text{var } \varepsilon'_{(n)} B_{nij} \varepsilon_{(n)} \rightarrow 0,$$

as  $n \rightarrow \infty$ . The former of these limits follows from Assumption 7; the latter follows from Assumption 6. ■

**Proof of Lemma 9.** From Proposition 1 we obtain

$$R_{nij}(\gamma) - R_{nij}(\gamma_o) = \beta_n(\gamma, \gamma_o) + b'_{(n)}(\gamma, \gamma_o) \varepsilon_{(n)} + \varepsilon'_{(n)} B_n(\gamma, \gamma_o) \varepsilon_{(n)},$$

where  $\varepsilon_{(n)} := y_{(n)} - \mu_{(n)}(\gamma_o)$  is distributed  $N(0, \Omega_n(\gamma_o))$ ,

$$\begin{aligned} \beta_n(\gamma, \gamma_o) := & - \left\{ \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma) \left( \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} \right) - \left( \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma_i} \right)' \Omega_n^{-1}(\gamma_o) \left( \frac{\partial \mu_{(n)}(\gamma_o)}{\partial \gamma_j} \right) \right\} \\ & - \frac{1}{2} \left\{ \text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \Omega_n(\gamma) \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \Omega_n(\gamma) \right) \right. \\ & - \text{tr} \left( \frac{\partial \Omega_n^{-1}(\gamma_o)}{\partial \gamma_i} \Omega_n(\gamma_o) \frac{\partial \Omega_n^{-1}(\gamma_o)}{\partial \gamma_j} \Omega_n(\gamma_o) \right) \left. \right\} + \frac{1}{2} \left\{ \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma) \right) \right. \\ & - \text{tr} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma_o)}{\partial \gamma_i \partial \gamma_j} \Omega_n(\gamma_o) \right) \left. \right\} - (\mu_{(n)}(\gamma) - \mu_{(n)}(\gamma_o))' q_{nij}(\gamma) \\ & - \frac{1}{2} (\mu_{(n)}(\gamma) - \mu_{(n)}(\gamma_o))' \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} (\mu_{(n)}(\gamma) - \mu_{(n)}(\gamma_o)), \end{aligned}$$

$$b_{(n)}(\gamma, \gamma_o) := \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} (\mu_{(n)}(\gamma) - \mu_{(n)}(\gamma_o)) + q_{nij}(\gamma) - q_{nij}(\gamma_o),$$

$$B_n(\gamma, \gamma_o) := -\frac{1}{2} \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \Omega_n^{-1}(\gamma_o)}{\partial \gamma_i \partial \gamma_j} \right)$$

and

$$q_{nij}(\gamma) := \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_i} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_j} + \frac{\partial \Omega_n^{-1}(\gamma)}{\partial \gamma_j} \frac{\partial \mu_{(n)}(\gamma)}{\partial \gamma_i} + \Omega_n^{-1}(\gamma) \frac{\partial^2 \mu_{(n)}(\gamma)}{\partial \gamma_i \partial \gamma_j}.$$

To complete the proof we need only show that conditions (i)–(v) of Lemma A.6 are satisfied. But this is easy in view of our assumptions: Taking  $A_n = I_n$ , (i) follows



from Assumption 9 (i)–(vi), (ii) follows from Assumption 9 (vii)–(ix), (iii) from Assumption 9 (x), and (iv) and (v) follow from Assumption 8. ■

**Proof of Theorem 3.** Assumption 2 follows from Assumption 2', Assumption 6 follows from Assumption 6' (iii), Assumptions 7(i) + (ii) follow from Assumptions 7'(ii) + (iii), and Assumptions 8(i) + (ii) follow from Assumption 6'(i) using the fact that  $((1/n) \text{tr } \Omega)^2 \leq (1/n) \text{tr } \Omega^2$ . Hence, if we can show that Assumption 9 holds, then Theorem 3 will follow from Theorem 2.

To prove conditions (i)–(x) of Assumption 9, let  $\phi \in \overset{o}{\Gamma}$  and  $\alpha > 0$  be arbitrary. Let  $N_o(\phi)$  be a bounded neighborhood of  $\phi$  such that Assumptions 6'–8' hold for all  $n \in \mathbb{N}$ ,  $\gamma \in N_o(\phi)$ , and  $\theta \in N_o(\phi)$ . Let  $N_1(\phi)$  be a "smaller" neighborhood of  $\phi$  such that its closure  $\bar{N}_1(\phi)$  is a subset of  $N_o(\phi)$ . Then

$$M_o := \sup \{M(\gamma): \gamma \in \bar{N}_1(\phi)\}$$

and

$$M_1 := \sup \{M_1(\gamma, \theta): (\gamma, \theta) \in \bar{N}_1(\phi) \times \bar{N}_1(\phi)\}$$

are finite nonnegative numbers, because  $M(\cdot)$  and  $M_1(\cdot, \cdot)$  are continuous functions on a compact set. Now let  $\delta$  be an appropriately chosen (small) positive number depending on  $\alpha$ ,  $M_o$ , and  $M_1$ . In view of Assumption 8'(iii) we may then choose a neighborhood  $N_2(\phi) \subset N_1(\phi)$  such that

$$\sup_{\gamma \in N_2(\phi)} \max_{1 \leq i \leq n} \left| \lambda_i \left( \frac{\partial^2 \Omega_n^{-1}(\gamma)}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 \Omega_n^{-1}(\phi)}{\partial \gamma_i \partial \gamma_j} \right) \right| \leq \delta, \quad (i, j = 1, \dots, p),$$

and

$$\sup_{\gamma \in N_2(\phi)} (\gamma - \phi)'(\gamma - \phi) \leq \delta.$$

We now claim that Assumption 9 holds for every  $n \in \mathbb{N}$  and  $\gamma \in N_2(\phi)$ . In particular,  $7'(\text{ii}) + 8'(\text{i}) \Rightarrow 9(\text{i})$ ,  $7'(\text{iii}) + 8'(\text{i}) \Rightarrow 9(\text{ii})$ ,  $6'(\text{ii}) + (\text{iii}) \Rightarrow 9(\text{iii})$ ,  $6'(\text{i}) + (\text{ii}) + (\text{iii}) + 8'(\text{iii}) \Rightarrow 9(\text{iv})$ ,  $7'(\text{ii}) + (\text{iii}) + 8'(\text{i}) \Rightarrow 9(\text{v})$ ,  $7'(\text{i}) + 8'(\text{ii}) \Rightarrow 9(\text{vi})$ ,  $8'(\text{ii}) \Rightarrow 9(\text{vii})$ ,  $7'(\text{iv}) + 8'(\text{ii}) \Rightarrow 9(\text{viii})$ ,  $2' + 7'(\text{iv}) + (\text{v}) \Rightarrow 9(\text{ix})$ , and  $8'(\text{iii}) \Rightarrow 9(\text{x})$ .

A detailed proof of these ten implications is tedious but straightforward, and is available from the authors upon request. Repeated use is made of Cauchy–Schwarz inequality, the mean-value theorem and the mean-value theorem for vector functions (Apostol [2, p. 355]). ■

**Proof of Theorem 5.** We shall verify the conditions of Theorem 4, i.e., Assumptions 1, 2', 3–5 and 6''–8''.

Assumption 1 follows from C1. Assumption 2' follows from C2 and the fact that  $\sigma^2 \rho^{t-1}/(1 - \rho^2)$  is a twice continuously differentiable function of  $\sigma^2$  and  $\rho$  for every  $t \in \mathbb{N}$ . To prove Assumption 3 we note that  $|V_n(\rho)| = 1/(1 - \rho^2)$ , so that  $\Omega_n(\sigma^2, \rho) := \sigma^2 V_n(\rho)$  is positive definite for every  $n \in \mathbb{N}$ ,  $\sigma^2 > 0$ ,  $|\rho| < 1$ .



To demonstrate Assumption 4 we define the  $n \times n$  matrices

$$E_n := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad \text{and} \quad \Delta_n := \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

The inverse of  $\Omega_n$  can now be expressed as

$$\Omega_n^{-1}(\sigma^2, \rho) = (1/\sigma^2)(I_n + \rho^2 \Delta_n - \rho(E_n + E'_n)),$$

from which we obtain

$$\partial \Omega_n^{-1}(\sigma^2, \rho) / \partial \sigma^2 = -(1/\sigma^2) \Omega_n^{-1}(\sigma^2, \rho),$$

and

$$\partial \Omega_n^{-1}(\sigma^2, \rho) / \partial \rho = (1/\sigma^2)(2\rho \Delta_n - (E_n + E'_n)).$$

Hence,

$$\frac{1}{2} \text{tr}((\partial \Omega_n^{-1} / \partial \sigma^2) \Omega_n)^2 = n/(2\sigma^4),$$

$$\frac{1}{2} \text{tr}((\partial \Omega_n^{-1} / \partial \rho) \Omega_n)^2 = (n-1)/(1-\rho^2) + 2\rho^2/(1-\rho^2)^2,$$

and

$$\frac{1}{2} \text{tr}((\partial \Omega_n^{-1} / \partial \sigma^2) \Omega_n (\partial \Omega_n^{-1} / \partial \rho) \Omega_n) = \rho/(\sigma^2(1-\rho^2)).^8$$

Now, with  $H_n(\beta, \rho) := (\partial \mu_{(n)}(\beta) / \partial \beta')' V_n^{-1}(\rho) (\partial \mu_{(n)}(\beta) / \partial \beta')$ , the symmetric  $(k+2) \times (k+2)$  matrix  $G_n(\beta, \rho, \sigma^2)$  takes the form

$$G_n(\beta, \rho, \sigma^2) = \begin{bmatrix} (1/\sigma^2)H_n(\beta, \rho) & 0 & 0 \\ 0' & \frac{n-1}{1-\rho^2} + \frac{2\rho^2}{(1-\rho^2)^2} & \frac{\rho}{\sigma^2(1-\rho^2)} \\ 0' & \frac{\rho}{\sigma^2(1-\rho^2)} & n/(2\sigma^4) \end{bmatrix},$$

and, using C3, the matrix  $(1/n) G_n(\beta, \rho, \sigma^2)$  converges as  $n \rightarrow \infty$ , to

$$G_o(\beta, \rho, \sigma^2) = \begin{bmatrix} (1/\sigma^2)H_o(\beta, \rho) & 0 & 0 \\ 0' & 1/(1-\rho^2) & 0 \\ 0' & 0 & 1/(2\sigma^4) \end{bmatrix}.$$



Furthermore, the matrix  $G_o(\beta, \rho, \sigma^2)$  is positive definite for every  $\beta \in \overset{o}{B}$ ,  $|\rho| < 1$ ,  $\sigma^2 > 0$ , because  $H_o(\beta, \rho)$  is positive definite.

To demonstrate Assumption 5 we define the  $n \times n$  matrices

$$Z_n(\rho) := I_n - \rho E_n, \quad \text{and} \quad A_n(\sigma^2, \rho) := \sigma^2(I_n + (\rho^2/(1 - \rho^2))e_{(n)}e'_{(n)}),$$

where  $e_{(n)}$  is the first column of  $I_n$ . Then  $\Omega_n^{-1} = Z_n A_n^{-1} Z_n'$  is the Cholesky decomposition of  $\Omega_n^{-1}$ , and

$$\begin{aligned} (1/n^2) \operatorname{tr}((\partial Z_n / \partial \rho) A_n^{-1} (\partial Z_n / \partial \rho)' \Omega_n)^2 \\ &= (1/n^2) \operatorname{tr}(E_n(I_n - \rho^2 e_{(n)} e'_{(n)}) E_n' V_n)^2 \\ &= (1/n^2) \operatorname{tr}(E_n' V_n E_n)^2 = (1/n^2) \operatorname{tr}(V_{n-1}^2) \\ &= (n(1 - \rho^2) - (1 - \rho^{2n})) / (n^2(1 - \rho^2)^4) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ .

Next, let us verify Assumptions 6''–8''. We shall need the following result. Let  $A = (a_{ij})$  be an arbitrary (possibly complex)  $n \times n$  matrix and let

$$R := \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \quad \text{and} \quad T := \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|.$$

Then

$$\max_{1 \leq t \leq n} |\lambda_t(A)| \leq \min(R, T).$$

Proof. See [14, p. 145]. ■

Using this result, and noting that

$$\partial^2 \Omega_n^{-1}(\sigma^2, \rho) / \partial \sigma^2 \partial \sigma^2 = (2/\sigma^6) V_n^{-1}(\rho), \tag{B.4}$$

$$\partial^2 \Omega_n^{-1}(\sigma^2, \rho) / \partial \rho \partial \rho = (2/\sigma^2) \Delta_n, \tag{B.5}$$

$$\partial^2 \Omega_n^{-1}(\sigma^2, \rho) / \partial \sigma^2 \partial \rho = -(1/\sigma^4)(2\rho \Delta_n - (E_n + E_n')), \tag{B.6}$$

we obtain

$$\max_{1 \leq t \leq n} \lambda_t(\Omega_n(\sigma^2, \rho)) \leq 2\sigma^2/(1 - |\rho|)^2$$

and

$$\max_{1 \leq t \leq n} |\lambda_t(Q_n^{(h)}(\sigma^2, \rho))| \leq 4(\sigma^4 + 2)/\sigma^6, \quad (h = 1, \dots, 6),$$



where

$$Q_n^{(1)} := \Omega_n^{-1}, \quad Q_n^{(2)} := \partial \Omega_n^{-1} / \partial \sigma^2, \quad Q_n^{(3)} := \partial \Omega_n^{-1} / \partial \rho,$$

$$Q_n^{(4)} := \partial^2 \Omega_n^{-1} / \partial \sigma^2 \partial \sigma^2, \quad Q_n^{(5)} := \partial^2 \Omega_n^{-1} / \partial \rho \partial \rho,$$

$$Q_n^{(6)} := \partial^2 \Omega_n^{-1} / \partial \sigma^2 \partial \rho.$$

This shows that Assumption 6'' holds. Assumption 7'' follows from C4.

Finally, to prove that Assumption 8'' holds, we note from (B.4)–(B.6) that

$$\begin{aligned} Q_n^{(4)}(\sigma^2, \rho) - Q_n^{(4)}(\sigma_o^2, \rho_o) \\ = 2(\sigma^2 \sigma_o^2)^{-3} ((\sigma_o^6 - \sigma^6) I_n + (\rho^2 \sigma_o^6 - \rho_o^2 \sigma^6) \Delta_n + (\rho_o \sigma^6 - \rho \sigma_o^6)(E_n + E'_n)), \end{aligned}$$

$$Q_n^{(5)}(\sigma^2, \rho) - Q_n^{(5)}(\sigma_o^2, \rho_o) = 2(\sigma^2 \sigma_o^2)^{-1} (\sigma_o^2 - \sigma^2) \Delta_n,$$

and

$$Q_n^{(6)}(\sigma^2, \rho) - Q_n^{(6)}(\sigma_o^2, \rho_o) = (\sigma^2 \sigma_o^2)^{-2} ((2\rho_o \sigma^4 - 2\rho \sigma_o^4) \Delta_n + (\sigma_o^4 - \sigma^4)(E_n + E'_n)).$$

Hence we obtain (after some elementary arithmetic) for  $h = 4, 5, 6$ ,

$$\begin{aligned} \max_{1 \leq t \leq n} |\lambda_t(Q_n^{(h)}(\sigma^2, \rho) - Q_n^{(h)}(\sigma_o^2, \rho_o))| \\ \leq (2(1 + \sigma^2)(1 + \sigma_o^2)/(\sigma^2 \sigma_o^2))^3 (|\sigma^2 - \sigma_o^2| + |\rho - \rho_o|). \end{aligned}$$

This concludes the proof of Theorem 5.